Conformational Changes in Ligand Binding Processes

Conformational Changes in Ligand Binding Processes

Dissertation
for the award of the degree
Doctor rerum naturalium
of the Georg-August-Universität Göttingen

within the doctoral program IMPRS-PBCS
of the Georg-August University School of Science (GAUSS)

Submitted by
Béla Voß
from Hamburg

Göttingen, 2014

- Thesis Committee:
  - Prof. Dr. Helmut Grubmüller
    Department for Theoretical and Computational Biophysics
    Max-Planck-Institute for Biophysical Chemistry
  - Prof. Dr. Jörg Enderlein
    Third Institute of Physics
    Faculty of Physics
    Georg-August-Universität Göttingen
  - Prof. Dr. Marina Benatti
    Electron Spin Resonance Spectroscopy Group
    Max-Planck-Institute for Biophysical Chemistry
- Examination Board:
  - Prof. Dr. Helmut Grubmüller (supervisor, reviewer)
    Department for Theoretical and Computational Biophysics
    Max-Planck-Institute for Biophysical Chemistry
  - Prof. Dr. Jörg Enderlein (reviewer)
    Third Institute of Physics
    Faculty of Physics
    Georg-August-Universität Göttingen
  - Prof. Dr. Marina Benatti
    Electron Spin Resonance Spectroscopy Group
    Max-Planck-Institute for Biophysical Chemistry
  - Prof. Dr. Sarah Köster
    CRC Research Group "Nanoscale Imaging of Cellular Dynamics"
    Faculty of Physics
    Georg-August-Universität Göttingen
  - Dr. Thomas Burg
    Biological Micro- and Nanotechnology Group
    Max-Planck-Institute for Biophysical Chemistry
  - Prof. Dr. Markus Müller
    Institute for Theoretical Physics
    Faculty of Physics
    Georg-August-Universität Göttingen

Date of the disputation: 2015-01-30

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Göttingen, den 26.11.2014

Béla Voß

Contents

List of Abbreviations **xi**

I. Opening 1

1. Introduction 3

2. Theoretical Background 5
2.1. Molecular Dynamics Simulations 5
2.1.1. From Quantum Mechanics to Molecular Dynamics - Approximations for Force Field Based Molecular Dynamics 5
2.1.2. Performing Molecular Dynamics Simulations 10
2.2. Enhanced Sampling via Umbrella Sampling 11
2.2.1. Potential of Mean Force 11
2.2.2. Umbrella Sampling 12
2.2.3. Weighted Histogram Analysis Method 14
2.2.4. Hamiltonian Replica Exchange 14
2.3. Principal Component Analysis 15
2.4. Rate Estimation 16
2.4.1. Rate Estimation if no Transitions Occur 19
2.5. Diffusion Controlled Kinetics 21
2.5.1. Rotational Restriction 22

II. The Ligand Binding Mechanism of MloK1 25

3. Introduction 27

4. The Conformational Change in the CNBD 31
4.1. Introduction 31
4.2. Theory 33
4.2.1. Derivation of a Reaction Coordinate for the Conformational Change 33
4.3. Methods 35
4.3.1. Molecular Dynamics Simulations 35
4.3.2. Data Analysis 38

4.4. Results and Discussion . . . 39
4.4.1. Derivation of a Reaction Coordinate . . . 39
4.4.2. Free Energy Landscapes . . . 41
4.4.3. Rates of Conformational Changes . . . 46
4.4.4. Rates of Unbinding . . . 48
4.5. Conclusion . . . 52

5. The Binding of cAMP **55**
5.1. Introduction . . . 55
5.2. Theory . . . 56
5.2.1. Substates and Pathways in a Ligand Binding Process . . . 56
5.2.2. Markov Model of Ligand Binding . . . 61
5.3. Methods . . . 62
5.3.1. Molecular Dynamics Simulations . . . 62
5.3.2. Data Analysis . . . 63
5.4. Results and Discussion . . . 66
5.4.1. Markov Model . . . 66
5.4.2. Protein Surface Attachment . . . 69
5.4.3. Three-State Model . . . 73
5.4.4. Binding Funnel . . . 75
5.5. Conclusion . . . 83

III. Long-range Allostery of CRM1 **85**

6. Introduction **87**
6.1. Own Contribution . . . 90

7. Structural Basis for Cooperativity of CRM1 Export Complex Formation **93**

8. Structural Determinants and Mechanism of Mammalian CRM1 Allostery **113**

9. Conclusion **125**

IV. Closing **129**

10. Summary and Outlook **131**

Appendix **133**

Bibliography **133**

Acknowledgements **141**

Acronyms

cAMP cyclic adenosine monophosphate

CNBD cyclic nucleotide binding domain

CNG cyclic nucleotide gated

COM centre of mass

CRM1 chromosome region maintenance 1

MD molecular dynamics

NES nuclear export signal

NMR nuclear magnetic resonance

NPC nuclear pore complex

PCA principle component analysis

PMF potential of mean force

RanGTP the GTPase Ran in its GTP-bound form

RMSD root mean square deviation

SPN1 snurportin1

WHAM weighted histogram analysis method

Part I.

Opening

1. Introduction

Ligand and protein-protein binding processes play a crucial role in biological processes. They are an essential part of the immune response; they are very often a first step in signalling processes, for example at G-protein coupled receptors; they are pre-requisites to ion channel activation, inactivation or modulation and they are an obvious part of cargo transport processes.

An important subclass are processes where the protein adopts different conformations depending on the presence or absence of a binding partner.

Since so many physiological processes are triggered, regulated or inhibited by ligand binding, many pharmaceutical drugs also work by binding to specific receptors. Therefore, the understanding of these processes may also assist for the development of new pharmaceutical drugs.

Consequently a lot of effort has been put into the investigation of such processes. Many experiments have been carried out to study the effects of ligand binding and the affinities of ligands to receptors. Furthermore, structural information on ligand binding processes has been obtained via x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy. In those cases where the three-dimensional atomistic structure of the unbound and the bound complex has been solved, these structures represent the "end states" of the binding process.

Much less is known on the pathways that connect these end states. Beside affinities, in most cases, the only available information are the rates between the unbound and bound state and vice versa, because it is usually difficult investigate the causal and temporal relation between ligand or protein binding and the conformational changes on an atomistic level.

One way to fill the gap are all-atom molecular dynamics (MD) simulations, which are a powerful tool to investigate the molecular motions and conformational changes in proteins at atomistic resolution. Using the structural information about the end states they are principally very well suited to study the detailed mechanism of binding processes.

In this thesis we want to address two major questions that are directly related to the pathway between the aforementioned end states: First: What is the sequential order of ligand binding and conformational change and what are the kinetics of the microscopic processes? Second: How does the binding lead to a conformational change and what are the functional consequences?

To address these questions, we selected two different systems where the bound and unbound conformation are known and where binding plays a different functional role: The binding of the cyclic nucleotide cyclic adenosine monophosphate (cAMP) at the cyclic nucleotide binding domain (CNBD) of the potassium channel

MloK1 from the bacterium *Mesorhizobium loti* and the allosteric binding of proteins at the karyopherin, i.e. transport protein, chromosome region maintenance 1 (CRM1).

In the former system, the CNBD of MloK1, binding at the binding domain leads to a conformational change in the binding domain that then influences the conductivity of the ion channel. The systems is highly suited to address our first questions because additionally to structural information on the binding site there are also kinetic informations, i.e. the global on- and off-rates, available. Furthermore mutational studies have been carried out that already suggest that binding occurs before the conformational change. The availability of all these experimental data enables us to compare our findings with experimental measurements.

In this work we focus solely on the binding domain for which we postulate a set of substates. We then develop and use a methodology to calculate the rates and free energy differences for transitions between these substates to understand the microscopic dynamics, determine the sequential order of binding and conformational change and thereby classify the mechanism and derive estimates for effective on and off-rates which are then compared to experimentally measured rates. It will be fully introduced and discussed in detail in part II.

In the latter system, CRM1, binding of a signal protein leads to a conformational change that changes the proteins affinity for cargo proteins. The exact cause of this allostery remained unclear though. Again it is not only the availability of structural information on the bound and unbound conformations that renders the system suitable for investigation, but also the information on the biological role, i.e. the transport of proteins from the nucleus into the cytoplasm, and the information that the allostery exists. Additionally, experiments have shown that a specific part of the protein is necessary for the allostery but were not able to fully explain this requirement. How the cooperative binding is achieved and what causes the conformational changes will be fully discussed in part III.

First though we will introduce the basic theoretical background of the essentials techniques used in this work.

2. Theoretical Background

In this chapter we will describe the essential established concepts that lay the foundation of our work. The method of MD simulations will be introduced, as well as some special simulation and analysis techniques.

2.1. Molecular Dynamics Simulations

Biomolecules such as proteins are complex and microscopical many particle systems. Over time many sophisticated experimental techniques have been developed to investigate structure, dynamics and function of such molecules. Another and today complimentary approach is the computational evaluation of the time evolution of the many particle system.

Atomistic systems can be simulated on different levels of accuracy. For each level different approximations are made, leading to different models. The level used for the investigations in this thesis are atomistic force field based MD simulations. MD in general describes a class of simulation, where the interaction between single particles is given by a potential energy landscape and the dynamics is calculated by calculating the forces on the every particle and integrating the equations of motion.

The foundations of MD have been described in many publications and books such as (Berendsen, 2007; Frenkel and Smit, 2001; van der Spool et al., 2005).

In the following sections we will follow these established paths and briefly discuss the different levels of accuracy as well as the respective approximations and limitations as well as their appropriateness to the systems investigated in this work.

2.1.1. From Quantum Mechanics to Molecular Dynamics - Approximations for Force Field Based Molecular Dynamics

2.1.1.1. Description of a Nuclei-Electrons System

Properties and interactions of matter at the lowest scales and highest energies are still (and will probably remain for a long time) topic of ongoing research. As a consequence there is no approximation-free bottom level to conceptually describe real atomic system.

For biomolecules in realistic conditions however extreme energies and temperatures do not play a role, nor do due to the small mass gravitational interactions. If

furthermore all interactions within the atomic nucleus are ignored, the basic level consists of nuclei and electrons with properties mass, charge and spin and solely electromagnetic interactions. The time evolution is then given by the laws of relativistic quantum mechanics. Already in this framework radioactive decay cannot be described or accounted for as it is a by-product of strong and weak interactions within the nucleus.

To calculate any time evolution for a realistic biomolecular system, several further approximations are necessary, which shall briefly be discussed in the following.

2.1.1.2. Nonrelativistic Conditions

As biomolecules consist mainly of light nuclei, the non-relativistic energy-momentum relation $E = p^2/2m$ holds even for the inner electrons, therefore a system of nuclei and electrons can be described by nonrelativistic quantum mechanics.

For very heavy nuclei this no longer holds in general, direct relativistic effects however only affect the inner electrons shells. The indirect effects on the outer electron shells due to contraction of inner orbitals can be accounted for by effective core potentials, rendering a relativistic treatment of these electrons unnecessary.

As a result the system is defined by the wave function

$$\phi(\mathbf{r}_{electrons}, \mathbf{r}_{nuclei}, t) \tag{2.1}$$

and the time evolution is given by the Schrödinger equation:

$$i\hbar\partial_t\phi(\mathbf{r}_e, \mathbf{r}_n, t) = \hat{H}(\mathbf{r}_e, \mathbf{r}_n, t)\phi(\mathbf{r}_e, \mathbf{r}_n, t) \tag{2.2}$$

In the systems investigated the Hamiltonian is time-independent. In this case the solution is given by a superposition of solutions to the stationary Schrödinger equation:

$$\phi(\mathbf{r}_e, \mathbf{r}_n, t) = \sum_n c_n e^{-iE_n t/\hbar}\psi(\mathbf{r}_e, \mathbf{r}_n). \tag{2.3}$$

This reduces the task to solving the stationary Schrödinger equation, and is a prerequisite of the Born-Oppenheimer approximation.

2.1.1.3. Born-Oppenheimer-Approximation

The nuclei are several thousand times heavier than the electrons. This allows for the approximation that the wave function separates into an electronic and nuclear part, where the electronic part depends only parametrically on the nuclear coordinates and the nuclear part only depends on the nuclear coordinates:

$$\phi(\mathbf{r}_{electrons}, \mathbf{r}_{nuclei}, t) = \xi(\mathbf{r}_{electrons}; \mathbf{r}_{nuclei}) \cdot \chi(\mathbf{r}_{nuclei}, t). \tag{2.4}$$

This separation is possible, because changes in the electronic part happen on a much shorter timescale than changes in nuclear wave function. Therefore it is assumed that the electronic wave function adapts instantaneously to any change in the nuclear configuration.

Accordingly, the Hamiltonian is split into an electronic part and a nuclear part. Under the assumption that the electronic part stays in one eigenstate, the energy eigenvalues of the electronic Hamiltonian for all possible nuclear configurations define an effective potential for the nuclei.

The Born-Oppenheimer approximation breaks if the energy eigenstates are (almost) degenerate or if the kinetic energy of the nuclei is so large that the timescales for the changes of the electronic and nuclear wave function do not separate anymore.

For the systems investigated in this thesis, it is safe to assume that the electronic wave function is always in the ground state and the nuclear kinetic energy is limited to room temperature kinetic energy. Therefore the Born-Oppenheimer approximation is very well justified.

Indeed it has been shown (Handy and Lee, 1996) that the effect of the Born-Oppenheimer approximation on molecules in their ground state is small.

2.1.1.4. Classical Approximation for Nuclei Motion

Despite the separation of nuclear and electronic wavefunction the numeric solution of the nuclear Schrödinger equation for many particles such as the proteins investigated in this thesis remains computationally infeasible with current computational resources. Therefore the next approximation is to treat the nuclei as classical particles. With this assumption the task of calculating the time evolution is reduced from solving the Schrödinger equation for the nuclear coordinates to simple Newtonian dynamics in a potential generated by the solution of the electronic wave equation.

The approximation is handwavingly justified by the fact that the thermal de-Broglie wavelength at room temperature for most atom types is small compared to the interatomic distances.

The potential of chemical bonds however is very steep. Due to the small atomic masses the oscillations are largely quantum mechanical and cannot be reproduced in a classical treatment of the nuclei. This effect is accounted for by constraining these fast degrees of freedoms. In this work out interest is mainly focussed on large-scale motions and conformational changes in proteins and exact nature of bond fluctuations is thus of little relevance.

2.1.1.5. Empirical Force Field Approximation

To this extent the molecular dynamics of the nuclei is given by Newtons equation of motion in a potential.

This potential of completely determined by the energy eigenvalues of the electronic wave function for all nuclear configurations. The eigenvalues and their derivatives with respect to the nuclear coordinates remains to be calculated quantum mechanically. This resulting type of molecular dynamics is called Born-Oppenheimer molecular dynamics.

Another approach is to find an approximation to the potential by analytical functions depending on the nuclear coordinates, thus eliminating the need to solve the electronic Schrödinger equation for many nuclear configurations.

Chemical experience shows that the full potential $U(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ (for N atoms) can be approximated by individual terms for the potential of chemical bonds, angles, electrostatic interactions and van-der-Waals interactions. This approximation leads to a potential written as a sum of terms depending only on atom pairs, triples or quadruples. The collection of all terms with all constant parameters is generally referred to as a *force field*. The shape of these terms will be discussed below.

This last approximation is clearly a severe approximation whose borders are a priori unclear, rendering validation with experimental observations a crucial task.

On the other hand due to the drastically increased computational efficiency of the method, many systems are now eligible to become subject of investigation, including those treated in this thesis.

2.1.1.6. Terms of an Empirical Force Field

There is no unique force field, i.e. collection of interaction terms and parameters; instead over time many force fields have been developed. Typical force fields, including the amber99sb force field (Hornak et al., 2006) used for the simulations in this thesis consist of terms for so called bonded interactions which depend on neighbouring atoms and for pairwise non-bonded interactions.

The amber force field potential energy consists of the following terms:

$$
\begin{aligned}
U(\mathbf{r}) = \quad & \underbrace{U_{\text{angles}} + U_{\text{dihedrals}} + U_{\text{improper dihedrals}}}_{\text{bonded interactions}} && (2.5)\\
& + \underbrace{U_{\text{Lennard-Jones}} + U_{\text{Coulomb}}}_{\text{nonbonded interactions}} && (2.6)\\
& && (2.7)\\
= \quad & \underbrace{\sum_i \frac{k_{\alpha_i}}{2}(\alpha_i - \alpha_{i,0})^2}_{\text{angle vibrations}} && (2.8)\\
& + \underbrace{\sum_i \frac{k_{d_i}}{2}(1 + \cos(n\phi_i - \phi_{i,0}))^2}_{\text{dihedral vibrations and rotations}} && (2.9)\\
& + \underbrace{\sum_i \frac{k_{\xi_i}}{2}(\xi_i - \xi_{i,0})^2}_{\text{in-plane vibrations}} && (2.10)\\
& + \underbrace{\sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 \cdot d_{ij}}}_{\text{pairwise electrostatic interactions}} && (2.11)\\
& + \sum_{i<j} 4\epsilon_{ij} \left[\underbrace{\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{12}}_{\text{Pauli repulsion}} - \underbrace{\left(\frac{\sigma_{ij}}{d_{ij}}\right)^{6}}_{\text{Van-der-Waals attraction}} \right] && (2.12)
\end{aligned}
$$

Bonded Interactions. Angles between three neighbouring, bonded atoms are known to fluctuate around an equilibrium value. This is accounted for by an harmonic angle potential. Dihedral angles between four consecutive, bonded atoms are described by periodic potential functions. Aromatic rings within molecules are known to have a planar topology. The potential enforcing this behaviour is modelled by so called improper dihedrals.

Bonds. In historical formulations typically an additional harmonic term for bond vibrations is included. As stated above, the resulting classical bond vibration is unrealistic: The potential is so steep that the dynamics at typical temperatures is mostly quantum mechanical. Since bond oscillations barely couple with other degrees of freedom and are rarely of interest in MD simulations, the simulation level is usually coarsened by constraining the bond distances to their equilibrium value. This is achieved by the LINKS algorithm (Hess et al., 1997).

Non-bonded Interactions. All atoms are assigned an effective partial charge, such that the resulting potential mimics the potential caused by the nuclei and the electron density determined by the exact solution of the electronic Schrödinger equation. These partial charges are used to describe pairwise long distance interactions between atoms.

The electrostatic interactions are long-ranged and cannot be cut off without introducing severe inaccuracies. In simulation systems with periodic boundary conditions an improved version of Ewald summation (Ewald, 1921), particle mesh Ewald (Darden et al., 1993) is employed. Ewald summation decomposes the summation of pairwise interactions in periodic systems into a real space part and into a Fourier space part. By evaluating the charges in the Fourier space on a grid, the efficiency of the calculation can be improved such that the overall scaling is $\mathcal{O}(n \log n)$ (with n being the number of particles).

Attractive Van-der-Waals interactions due to spontaneously induced dipoles are known to scale with r^{-6}. For computational efficiency, the Pauli repulsion is calculated via an r^{-12}-term. Taken together these terms constitute the pairwise Lennard-Jones potential. It is fully determined by the position of its minimum and its value at the minimum. These parameters are usually determined for single atom types. To describe the interactions between atoms of different type, geometric or arithmetic averages are used.

2.1.2. Performing Molecular Dynamics Simulations

2.1.2.1. Integrating the Equations of Motion

In MD simulations the potential energy term is used to calculate the forces

$$\mathbf{F_i}(\mathbf{r}_1 \dots \mathbf{r}_N) = \frac{\partial U(\mathbf{r}_1, \dots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \tag{2.13}$$

on every atom. Subsequently Newtons equation of motion

$$\mathbf{F}(\mathbf{x}) = \mathbf{a} \cdot m = \ddot{\mathbf{x}} \cdot m \tag{2.14}$$

are integrated using the so called Verlet algorithm (Verlet, 1967) [1]:

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \frac{F(t)}{m} \Delta t^2. \tag{2.15}$$

Equivalent versions of the Verlet algorithm are the velocity Verlet algorithm (Swope et al., 1982) and the leapfrog method which have the advantage of directly providing accurate velocities.

The error per step is $\mathcal{O}(\Delta t^4)$. The global error in position for multiple timesteps

[1] The algorithm itself was used long before Verlet by Størmer and by Cowel and Crommelin and is therefore sometimes referred to as the Størmer method.

however is $\mathcal{O}(\Delta t^2)$.

Due to Lyapunov instability two trajectories with neighbouring starting parameters diverge very quickly. For the calculation of thermodynamic quantities this is however of little relevance: As long as overall energy conversation is ensured and assuming identity of the time and ensemble average the sampled probability distribution will remain correct.

2.1.2.2. Simulating Thermodynamics Ensembles Using Thermostats

Integration of the equations of motion yields an ensemble with constant energy, volume, particle number and overall momentum. This corresponds to the microcanonical ensemble known from statistical mechanics. Usual physiological and experimental conditions however are best described by the Gibbs ensemble with constant temperature, pressure and particle number.

To mimic the coupling of the simulation system to a heat and pressure bath, thermostat algorithms are used which rescale particle velocities and particle coordinates. In this work the velocity rescaling thermostat (Bussi et al., 2007), based on the Berendsen thermostat (Berendsen et al., 1984) and the Berendsen pressure coupling algorithm (Berendsen et al., 1984) are used for coupling of temperature and pressure, respectively.

2.2. Enhanced Sampling via Umbrella Sampling

2.2.1. Potential of Mean Force

The Helmholtz free energy of a system in n dimensions with Hamiltonian H is given by the logarithm of the partition function:

$$F = -kT \ln \int \mathrm{d}q^n \; e^{-\beta H(\mathbf{q})} \tag{2.16}$$

We define a reaction coordinate as a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ with non-vanishing gradient. The one-dimensional potential of mean force (PMF) along the reaction coordinate z is then defined as

$$W(z) = -kT \ln \int \mathrm{d}q^n \; \delta(f(\mathbf{q}) - z) \; e^{-\beta H(\mathbf{q})}. \tag{2.17}$$

The extension to multiple reaction coordinates is straightforward.

The value of the PMF is the free energy of the system if it is constraint along the reaction coordinate(s) to a specific value. The full PMF is therefore also denoted as the free energy landscape. The name potential of mean force plays tribute to the fact that the average force at a certain point along the reaction coordinates is given by the derivative of the PMF: $\langle F(z) \rangle = -\partial_z W(z)$.

The PMF is directly related to the equilibrium probability distribution along the reaction coordinate via the Boltzmann relation:

$$\rho(z) = \frac{e^{-\beta W(z)}}{\int \mathrm{d}z \; e^{-\beta W(z)}} \tag{2.18}$$

By inverting the relation the PMF can be calculated if the equilibrium distribution is known:

$$W(z) = -kT \ln \left(\frac{\rho(z)}{\rho(z_0)} \right) + W(z_0). \tag{2.19}$$

Since the transformation $W'(z) = W(z) + W_0$ does not change $\rho(z)$, $W(z)$ is only fixed up to a constant $W(z_0) + kT \ln \rho(z_0)$.

If a protein occupies multiple conformations which can be separated using one or a few coordinates, the PMF as the free energy landscape provides the basis to calculate free energy differences between these conformations and the barriers between them.

Therefore the calculation of PMFs is of high interest. The most straightforward technique is to calculate the equilibrium probability along the coordinate. With regular MD simulations this is often not computationally feasible, raising the need for sampling enhancing techniques which shall be discussed in the following section.

2.2.2. Umbrella Sampling

Equilibrium sampling techniques such as MD (or Monte Carlo for that matter) are techniques to calculate the probability function

$$p(q) \propto e^{-\beta U(q)} \tag{2.20}$$

with the potential energy U by directly counting of often a configuration q is sampled. The obvious problem is that in finite simulation time unlikely states are rarely or even never sampled leading to large errors in the estimate of their probability. For the calculation of free energy barriers this is a critical problem.

A possible solution is to introduce an additive biasing potential $V(q)$ that is small in regions of configurational space where $U(q)$ is large. This technique is commonly referred to as umbrella sampling (Torrie and Valleau, 1977). The biasing potential gives rise to biasing factor

$$w(q) = \exp(-\beta V(q)). \tag{2.21}$$

Although the sampled ensemble differs from the original ensemble, all ensemble averages from the original ensemble can be obtained. For an observable $A(q)$ it

holds:

$$\langle A \rangle = \frac{\int \mathrm{d}q \; A(q) \; e^{-\beta U(q)}}{\int \mathrm{d}q \; e^{-\beta U(q)}} \tag{2.22}$$

$$= \frac{\int \mathrm{d}q \; \frac{A(q)}{w(q)} \; w(q) \; e^{-\beta U(q)}}{\int \mathrm{d}q \; \frac{1}{w(q)} \; w(q) \; e^{-\beta U(q)}} \tag{2.23}$$

$$= \frac{\frac{\int \mathrm{d}q \; \frac{A(q)}{w(q)} \; w(q) \; e^{-\beta U(q)}}{\int \mathrm{d}q \; w(q) \; e^{-\beta U(q)}}}{\frac{\int \mathrm{d}q \; \frac{1}{w(q)} \; w(q) \; e^{-\beta U(q)}}{\int \mathrm{d}q \; w(q) \; e^{-\beta U(q)}}} \tag{2.24}$$

$$= \frac{\langle A/w \rangle_w}{\langle 1/w \rangle_w} \tag{2.25}$$

Here $\langle \cdot \rangle_w$ donates an average over the w-weighted ensemble. The average of $A(q)$, $\langle A \rangle$ is recovered by simply sampling in the weighted ensemble and averaging in this ensemble over $A(q)/w(q)$ and unbiasing by dividing through the inverse average weight.

If only probabilities (or free energies, i.e. PMFs) are of interest it holds:

$$\rho_{\text{bias}}(q) = \frac{w(q) \cdot e^{-\beta U(q)}}{\int \mathrm{d}q \; w(q) \; e^{-\beta U(q)}} \tag{2.26}$$

$$= \frac{w(q) \cdot \rho_{\text{unbiased}}(q) \cdot \int \mathrm{d}q \; e^{-\beta U(q)}}{\int \mathrm{d}q \; w(q) \; e^{-\beta U(q)}} \tag{2.27}$$

$$= \frac{w(q) \cdot \rho_{\text{unbiased}}(q)}{\langle w \rangle_0} \tag{2.28}$$

$$\Rightarrow \rho_{\text{unbiased}}(q) = w^{-1}(q) \cdot \rho_{\text{bias}}(q) \cdot \langle w \rangle_0 \tag{2.29}$$

For the PMF this leads to:

$$\Rightarrow W(q) \overset{(2.19,2.21)}{=} -kT \ln \left(\frac{\rho_{bias}(q)}{\rho(q_0)} \right) + W(q_0) + V(q) \underbrace{-kT \ln \left(\langle e^{-\beta V(q)} \rangle_0 \right)}_{F} \tag{2.30}$$

The last term, F, is a constant. It is the free energy associated with the introduction of the biasing potential. In the case of a single biasing potential $V(q)$ it can be absorbed into $W(q_0)$ and $\ln(\rho(q_0))$.

If the goal of the calculation is the reconstruction of a PMF, the use of a single biasing potential is often problematic as it requires already a good a-priori knowledge about $W(q)$. A remedy to this is to use multiple harmonic biasing umbrella potentials $V_i(q) = \frac{k}{2}(q - q_{i,0})^2$. For every umbrella window, the PMF is reconstructed via equation (2.30). However, the F_i are not identical for all umbrella windows and therefore have to be chosen consistently. Different approaches have been proposed to solve the problem, with the weighted histogram analysis method (WHAM) (Kumar et al., 1992) being the most state-of-the-art technique that is

also employed in this work.

2.2.3. Weighted Histogram Analysis Method

The WHAM (Kumar et al., 1992) expresses the total probability distribution as a weighted sum of the unbiased probability distributions $\rho_{i,\text{unbiased}}(q)$ of the individual umbrella windows.

This leads to an iterative approach: An estimate for the total unbiased probability distribution is calculated and subsequently the $F_i \propto \langle exp(-\beta V_i(q))\rangle_0$ are calculated via equation (2.30) using the total probability distribution. This estimate for the F_i is plugged in again to calculate a refined version of $\rho(q)$ and so forth. For well behaved systems and sufficient sampling this approach converges rapidly.

2.2.4. Hamiltonian Replica Exchange

The efficiency in umbrella sampling simulations is further increased when combined with parallel tempering, also known as replica exchange (Swendsen and Wang, 1986; Geyer, 1991; Earl and Deem, 2005). This combination is also employed in this work.

The idea of replica exchange is the following: Instead of simulating a single ensemble with fixed Hamiltonian and temperature and thus a single Boltzmann factor $\exp\left(-\frac{H(x)}{k_B T}\right)$, multiple replicas are simulated. In canonical replica exchange simulations the temperature differs between the replicas, yielding different Boltzmann factor. In Hamiltonian replica exchange simulations different Hamiltonians $H_i(x)$ are employed. The use of different Hamiltonians (or temperatures) leads to different sampled ensembles.

If the difference in the Boltzmann factors is small enough, configurations sampled in one replica have a non vanishing probability to be sampled in an other replica as well (and vice versa). in replica exchange simulations every replica is evolved according to its Hamiltonian. Increased sampling is reached by performing in regular intervals pairwise exchange attempts between the configurations of the replicas. To ensure that the ensemble in every replica still belongs to the Boltzmann distribution of the corresponding Hamiltonian (or temperature), the exchange probabilities are given by a Metropolis criterion. With the effective Hamiltonian $H_i^* = H_i(x)/K_b T$ and configurations q_i, q_j this probability is:

$$p(i \leftrightarrow j) = p_i(H_i^* \to H_j^*) \cdot p_j(H_j^* \to H_i^*) \tag{2.31}$$

$$= \min\left[1, \exp\left(H_i^*(q_i) + H_j^*(q_j) - H_i^*(q_j) - H_j^*(q_i)\right)\right]. \tag{2.32}$$

In umbrella sampling simulations the replicas correspond to the umbrella windows and the Hamiltonians differ by the different biasing potentials. The sampling advantage stems from the fact that barriers orthogonal to the sampling coordinate

that cannot be overcome in one replica/umbrella window may be overcome in another one. If therefore a configuration "walks" through different replicas it may thus pass the barrier before returning to its original replica. Therefore sampling in orthogonal degrees of freedom is enhanced.

The replica exchange step constitutes a Monte Carlo sampling step. Due to these steps, the trajectories no longer contain real time information. In umbrella sampling simulations however this does not matter, since only the sampled ensemble is of interest.

2.3. Principal Component Analysis

In the previous section we laid out how free energy differences and barriers between different protein conformations are calculated. This is achieved via the calculation of the PMF for reaction coordinates describing the relevant conformational change. A priori however these reaction coordinates are unknown and hard to identify simply due to the sheer amount of degrees of freedom in a protein.

Principle component analysis (PCA) is an orthogonal transformation of basis vectors, that allows for a given data set to identify the vectors with the largest variance. Applied to the ensembles of protein configuration it allows us to identify the largest conformational changes which are assumed to be most relevant for protein function. The coordinates describing large scale conformational changes are then used as the basis to identify suitable reaction coordinates. For this reason this section serves as a brief introduction to the basic concept of PCA.

The configurational space of a many particle system such as proteins is spanned by its $3N$ particle coordinates and every configuration constitutes a point in this $3N$-dimensional space. An ensemble of structures corresponds to a set of points, a time-continuous trajectory corresponds to a path in the $3N$-dimensional configurational space. Although proteins have many degrees of freedom, some of them, such as bonds or angles are constrained, meaning that the oscillations in these degrees of freedoms have small amplitudes. Additionally the interactions between the protein atoms enforce that the atoms move collectively. As any set of basis vectors spans the configurational space, one tries to find a basis where fewer basis vector suffice to describe as much of the protein dynamics as possible. PCA does this by identifying uncorrelated and orthogonal degrees of freedom with maximal variances.

The covariance between N random variables X_i is given by

$$\mathrm{Cov}(X_i, X_j) = \langle (X_i - \langle X_i \rangle) \cdot (X_j - \langle X_j \rangle) \rangle. \tag{2.33}$$

The elements of the covariance matrix are given by

$$c_{ij} = \mathrm{Cov}(X_i, X_j) \tag{2.34}$$

The covariance matrix is symmetric and according to the spectral theorem, it can

be diagonalised and the eigenvectors form an orthogonal basis:

$$\mathbf{C} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathbf{T}} \tag{2.35}$$

with a diagonal matrix $\mathbf{D}$ and an orthogonal matrix $\mathbf{Q}$. The eigenvalues $\lambda_i = d_{ii}$ are the variances along the transformed random variables $\mathbf{Y} = \mathbf{Q}^{\mathbf{T}}\mathbf{X}$. These transformed variables are by definition uncorrelated.

For ensembles of protein configurations, the average is taken over the individual configurations, for trajectories, i.e. time-sorted configurations, this corresponds to an average over time. The eigenvectors of the covariance matrix correspond to collective coordinates of all protein atoms. In case of trajectories, the eigenvector corresponding to the largest eigenvalue thus describes the collective coordinate where the largest motion occurs.

It has been shown that the eigenvalue spectrum of simulated protein trajectories decreases very rapidly with increasing eigenvalue numbers. This means that despite the high dimensionality of their configurational space most of the protein dynamics occurs in a much lower dimensional space. Assuming that these degrees of freedom are also the relevant one, this is an a posteriori justification that PCA is a useful method for protein dynamics analysis.

2.4. Rate Estimation

In our analysis of conformational changes of proteins we are not only interested on the free energy differences between the conformations, but also the probabilities of transitions between the states.

If a system's conformational space separates into states such that the timescales of mixing within the states and of transitions between the states separate, it is justified to approximate the kinetics between the states by rates. This means that the transitions are treated as Poisson processes and the probability to remain in one state for the time t is given by

$$p(t) = k \cdot e^{-kt} \tag{2.36}$$

with the rate k.

In this section we will derive how rates are estimated from equilibrium MD simulations.

Estimation of rates from MD simulations is rather straightforward: Simulations are started from configurations belonging to one state. The time evolution is monitored and if a transition into another state occurs, the point in time is recorded (figure 2.1).

The rate approximation is equivalent to treating the transition events as a Pois-

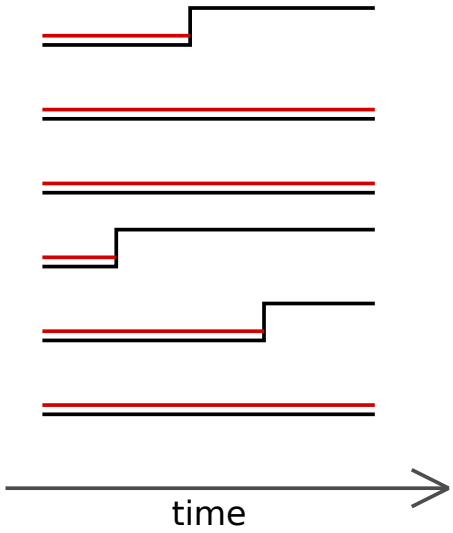


Figure 2.1.: Sketch of six timelines where a transition occurs (1st, 4th, 5th) or does not occur (2nd, 3rd, 6th). Black lines represent the full time lines. The cummulative length of the red lines is the cummulative time spent in the starting state.

son process. Subsequently the probability to observe n transitions is given by

$$p(n;\lambda) = \frac{\lambda^n}{n!} e^{-\lambda} \tag{2.37}$$

where the exception value λ is given by

$$\lambda = k \cdot T \tag{2.38}$$

with the "real" rate constant k and the total time T spent in starting state. This is the cummulative simulation time of the trajectories where no transitions occur and the sum of times until the transition for trajectories where a transition occurs (see figure 2.2). This leads to

$$p(n;k,T) = \frac{(kT)^n}{n!} e^{-kT}. \tag{2.39}$$

Using Bayes theorem to estimate the probability that a certain rate governs a process given a set of trajectories with a total of n transitions gives

$$p(k;n,T) = \frac{p(n;k,T) \cdot p(k)}{p(n;T)}. \tag{2.40}$$

The prior distribution for k is of course unknown. A reasonable estimate is that all possible time*scales* a equally likely. This corresponds to a probability

distribution

$$p(\log(k)) = \begin{cases} \frac{1}{b-a} & \forall\ \log(k) \in (a,b) \qquad a,b = const. \\ 0 & \text{else} \end{cases} \tag{2.41}$$

The lower and upper cutoff a and b ensure normalisability and can be chosen arbitrarily small and large. If rates are due to energetic barriers $k \propto exp(-\beta\Delta G^{\dagger})$ this corresponds to a uniform a-priori distribution of barrier heights.

The resulting probability distribution for k is then

$$p(k) = \frac{\mathrm{d}\log(k)}{\mathrm{d}k}\frac{1}{b-a} \ \forall\ \log(k) \in (a,b) \tag{2.42}$$

$$= \begin{cases} (k(b-a))^{-1} & \forall\ k \in (e^a, e^b) \\ 0 & \text{else} \end{cases} \tag{2.43}$$

If $n \geq 1$, the normalisation factor $p(n;T)$ is given by

$$p(n;T) = \int \mathrm{d}k\ p(n;k,T) \cdot p(k) \tag{2.44}$$

$$\approx \frac{T^n}{(b-a)n!} \int \mathrm{d}k\ k^{n-1}e^{-kT} \tag{2.45}$$

$$\overset{k=T/l}{=} \frac{1}{(b-a)n!} \int \mathrm{d}l\ l^{n-1}e^{-l} \tag{2.46}$$

$$= \frac{1}{(b-a)n} \tag{2.47}$$

Putting equations (2.39, 2.43 and 2.44) into equation (2.40) yields

$$p(k;n,T) = \frac{T^n}{(n-1)!}k^{n-1}e^{-kT} \tag{2.48}$$

Transformation to the average time $\tau = 1/k$ yields

$$p(\tau;n,T) = \frac{T^n}{(n-1)! \cdot \tau^{n+1}} \exp(-T/\tau) \tag{2.49}$$

For large n, equation (2.48 and 2.49) approach a Gaussian distribution.

2.4.0.1. Moments of k and τ

The moments of k are given by

$$\langle k^m \rangle = \frac{(n+m-1)!}{T^m (n-1)!} \tag{2.50}$$

$$\langle k \rangle = \frac{n}{T} \tag{2.51}$$

$$\mathrm{Var}(k) = \langle k^2 \rangle - \langle k \rangle^2 = \frac{n}{T^2} \tag{2.52}$$

The moments of τ are given by

$$\langle \tau^m \rangle = \frac{T^n (n-m-1)!}{(n-1)!} \tag{2.53}$$

$$\langle \tau \rangle = \frac{T}{n-1} \tag{2.54}$$

$$\mathrm{Var}(\tau) = \langle \tau^2 \rangle - \langle \tau \rangle^2 = \frac{T^2}{(n-1)^2 (n-2)} = \frac{\langle \tau \rangle^2}{(n-2)} \tag{2.55}$$

2.4.1. Rate Estimation if no Transitions Occur

If no transitions are observed i.e. $n = 0$, the normalisation integral (2.44) depends on the exact choice of c. Obviously an exact rate determination is impossible, however, if the rate is determined by an energetic barrier, a lower estimate for the height of the barrier is possible.

The normalisation in equation (2.40) becomes

$$p(n=0;T) = \int \mathrm{d}k \; p(0;k,T) \cdot p(k) \tag{2.56}$$

$$= \int_{e^a}^{e^b} \mathrm{d}k \; \frac{e^{-kT}}{k(b-a)} \tag{2.57}$$

$$= \frac{E_1(e^a T) - E_1(e^b T)}{b-a} \tag{2.58}$$

$$\approx \frac{E_1(e^a T)}{b-a} \tag{2.59}$$

with the exponential integral $E_1(x) = \int \mathrm{d}t \; \exp(-tx)/t$.

Thus it follows:

$$p(k; n=0, T) = \frac{e^{-kT}}{k \cdot E_1(e^a T)} \; \forall \; k > e^a \tag{2.60}$$

which still depends critically on e^a, which is basically the "lowest imaginable rate".

To motivate how are a lower boundary for the rate is estimated we assume the

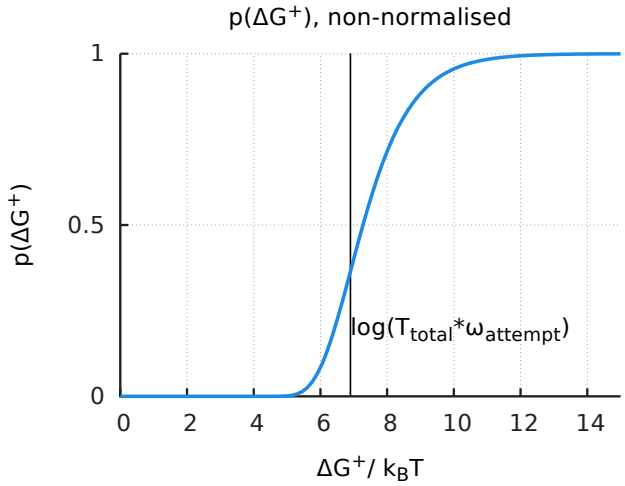


Figure 2.2.: Sketch of $p(\Delta G^{\dagger}; n = 0)$ for $\omega \cdot T = 1000$. The probability density changes within a small band from almost 0 to a constant value.

rate to be given by an energy barrier $\Delta G^{\dagger}$ and an attempt frequency ω via

$$k = \omega \cdot e^{-\beta \Delta G^{\dagger}}. \tag{2.61}$$

Expressing equation (2.60) in terms of $\Delta G^{\dagger}$ yields

$$p(\Delta G^{\dagger}) = \frac{\beta \cdot \exp(-T\omega e^{-\beta \Delta G})}{E_1(e^a)} \; \forall \; \Delta G^{\dagger} < \frac{1}{\beta} \ln\left(\frac{e^a}{\omega}\right). \tag{2.62}$$

The absolute value of equation (2.62) still depends on a, but the shape does not (see figure 2.2): It is sigmoidal, being almost 0 for small values of $\Delta G^{\dagger}$ and then within a rather small range changes to a maximum value. If knowledge about ω can be obtained otherwise the turning point can used as an estimate for the lower boundary of $\Delta G^{\dagger}$.

The turning point is given by:

$$\frac{\mathrm{d}^2 p(\Delta G^{\dagger})}{\mathrm{d}\Delta G^2} \stackrel{!}{=} 0 \tag{2.63}$$

$$\Delta G^* = \frac{\ln(T\omega)}{\beta} \tag{2.64}$$

This corresponds to a rate of

$$k > \frac{1}{T}. \tag{2.65}$$

Without more detailed knowledge about the lowest possible rate this is a reason-

able conservative estimate: If for example the lowest imaginable rate is two orders of magnitude smaller than $1/T$, i.e.

$$k_{low} = e^a = \frac{1}{100T} \tag{2.66}$$

then

$$p\left(k > \frac{1}{T_0}; T = T_0; n = 0; k_{low} = \frac{1}{100T_0}\right) = 0.054 \tag{2.67}$$

which is reasonably small.

2.5. Diffusion Controlled Kinetics

Ligand binding processes consist at least partially of a diffusion type part, where the two molecules diffuse in solution until they encounter. This also holds for the ligand binding processes discussed in this thesis. In this chapter we will briefly describe the very basic theoretical results for diffusion kinetics, based on the ground-laying work of Smoluchowski (1917).

Smoluchowski calculated the rate for the encounter of two particles by solving the radial diffusion equation with the boundary conditions that the concentration is constant at $t = 0$ for a volume outside a target sphere with radius R and 0 at the boundary of the target sphere for $t > 0$.

The radial diffusion equation is given by

$$\frac{\partial}{\partial t}\left(r \cdot c(r,t)\right) = D\frac{\partial^2}{\partial r^2}(r \cdot c(r,t)) \tag{2.68}$$

for the concentration c and with the diffusion coefficient D. The boundary conditions are

$$c = \begin{cases} c_0 & \text{for } t = 0,\ r > R \\ 0 & \text{for } t > 0,\ r = R. \end{cases} \tag{2.69}$$

The solution to this differential equation is

$$u = c_0\left(1 - \frac{R}{r} + \frac{2R}{r\sqrt{\pi}}\int_0^{\frac{r-R}{2\sqrt{Dt}}} e^{-x^2}\mathrm{d}x\right). \tag{2.70}$$

The flow through the sphere is the given by

$$J(t) = 4\pi R^2 D\frac{\partial c(r,t)}{\partial r}\Big|_{r=R} = 4\pi R D c_0\left(1 + \frac{R}{\sqrt{\pi Dt}}\right). \tag{2.71}$$

For sufficiently small times t the probability that a single particle in the volume V

reaches R is then

$$p = \frac{4\pi RD}{V}\left(t + \frac{2R\sqrt{t}}{\sqrt{\pi D}}\right). \tag{2.72}$$

Thus the probability that none of n particles reaches the target volume is given by

$$u(t) = (1-p)^n \overset{n\to\infty, p\to 0}{\longrightarrow} \exp(-p) \tag{2.73}$$

$$= \exp\left[-\underbrace{4\pi RDc_0 \cdot \left(1 + \frac{2R}{\sqrt{\pi Dt}}\right)}_{k_{\text{on}}} \cdot t\right]. \tag{2.74}$$

Comparing this term with $u(t) = \exp(-k_{on} \cdot t)$ yields the effective on-rate. The second term $\frac{2R}{\sqrt{\pi Dt}}$ decays with increasing t, for long timescales the rate constant is therefore given by

$$k = 4\pi N_A DR \tag{2.75}$$

which is the classical Smoluchowski rate.

For small molecules, the diffusion coefficient is of the order of $D \approx 10^{-9}\,\mathrm{m^2/s}$. If the target radius is also in the nm range the second term is close to 1 for processes on the ns scale. For longer timescales, i.e. 1 µs and longer, it becomes irrelevant. For smaller timescales though the apparent on-rate is faster than the classical Smoluchowski rate and thus not negligible.

If both ligand and receptor are free to diffuse, which is the usually case, the diffusion coefficient has to be replaced by the sum of the individual diffusion coefficients. If however the receptor is much larger than the ligand then due to stokes friction only the diffusion coefficient of the ligand is of importance.

2.5.1. Rotational Restriction

If molecules do no bind uniformly, but only at smaller sized patches, the effective on-rate decreases. Šolc and Stockmayer (1971, 1973) calculated analytical and approximate solutions for cases of spherical and axial symmetric molecules.

For the sake of brevity we will discuss one important result only. First we will not distinguish between receptor and ligand. Second the simpler case shall be considered where one molecule binds isotropically on its surface while the other only at a small circular patch with surface fraction size Φ_A. The average time for rotations between fitting and non-fitting orientation is τ_A. In this case the effective on-rate is given by

$$k_{\text{on, eff}} = \frac{k_{\text{Smoluchowski}} \cdot \Phi_A}{\frac{\Phi_A + k_{\text{off}} \cdot \tau_A}{1 + k_{\text{off}} \cdot \tau_A}}. \tag{2.76}$$

The equation has two interesting limiting cases:

$$k_{\text{on, eff}} = \begin{cases} k_{\text{Smoluchowski}} \cdot \Phi_A & \text{for } k_{\text{off}} \cdot \tau_A \gg \Phi_A \\ k_{\text{Smoluchowski}} & \text{for } k_{\text{off}} \cdot \tau_A \ll \Phi_A. \end{cases} \tag{2.77}$$

For slow rotational diffusion the Smoluchowski rate is reduced by the surface factor, for fast rotational diffusion however it is not influenced at all.

Part II.

The Ligand Binding Mechanism of MloK1

3. Introduction

Ligand-protein binding processes play an essential role in biological systems, be it immune response (Springer, 1990; Carreno and Collins, 2002), enzymatic activity or signalling (Kristiansen, 2004; Kaupp and Seifert, 2002). An import class of ligands occurring in signalling processes are cyclic nucleotides such as cAMP that again play a role in various physiological contexts (Pastan et al., 1975; Beavo and Brunton, 2002; Newton and Smith, 2004; Kaupp and Seifert, 2002). cAMP binding to proteins occurs at a cyclic nucleotide binding domain (CNBD). An important class here are cyclic nucleotide gated (CNG) channels (Kaupp and Seifert, 2002; Cukkemane et al., 2011) which are modulated by the presence or absence of cyclic nucleotides.

As the expression and crystallisation of eukaryotic CNG channels is difficult, endeavours have focused in the past on procaryotic CNG channels. An experimentally well studied example is the potassium channel MloK1 found in the bacterium *Mesorhizobium loti* and its CNBD (Nimigean et al., 2004; Nimigean and Pagel, 2007; Chiu et al., 2007; Cukkemane et al., 2007; Kowal et al., 2014). The conductivity of the ion channel has been shown to be increased in the ligand-bound conformation (Clayton et al., 2004).

For this system various x-ray (Clayton et al., 2004; Altieri et al., 2008) and NMR (Schünke et al., 2009, 2011) structures have been solved in the past. The structures reveal that ligand binding is associated with conformational changes in the CNBD: While the binding domain adopts an open conformation in the absence of a ligand, it assumes a closed conformation in presence of a ligand. (Here, *open* and *closed* relate to the visual impression of the CNBD. For the ion channel the situation is reversed: If the ligand is bound, the channel is conducting, therefore open. If the ligand is not bound, the ion channel is not conducting, i.e. closed.) The conformational change is mainly given by the rearrangement of the N-terminal helices, which constitute the connection to the transmembrane part of the channel, and the rearrangement of the C-terminal helix that closes like a lid over the binding pocket. The conformations are sketched in figure 3.1. In addition to the structural information, experimental data on stop-flow experiments that measure the effective on- and off-rate is available.

The availability of the structural information as well as the kinetic measurements render the system an exceptionally suitable system to computationally investigate the interplay between ligand binding and conformational change on an atomistic level.

Two concepts have been established to describe this interplay in ligand binding processes that are associated with a conformational change in the receptor: Induced

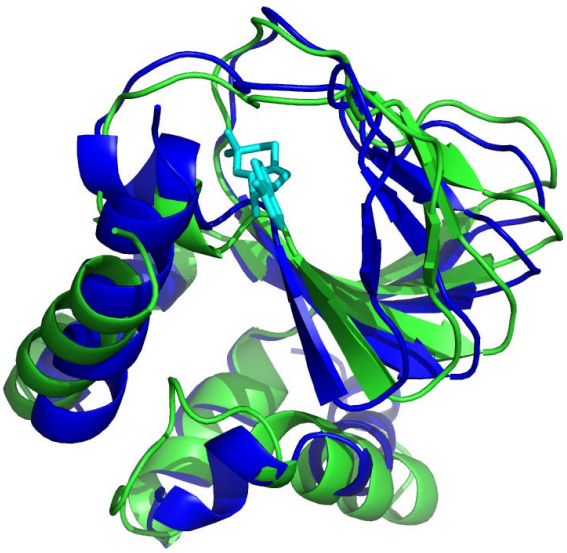


Figure 3.1.: X-ray structures of the open ligand-free (green) and closed ligand-bound (blue, with cyan coloured ligand) CNBD conformation.

fit and conformational selection (Csermely et al., 2010). In the induced fit model, the ligand first binds to the protein before any conformational change occurs. In the conformational selection model, on the other hand, both protein conformations are accessible even in the absence of the ligand. Only upon ligand binding the ligand-bound conformation, in case of the CNBD of MloK1 the closed conformation, becomes the preferred one.

A way to classify the binding process and the associated conformational changes is to determine the free energy differences and rates between the different conformations as well as the ligand-bound and unbound states. Based on these microscopic kinetics we then deduced the effective on- and off-rate between the open ligand-free conformation and the closed ligand-bound conformation, to allow the comparison with experimentally measured on- and off-rates.

How can we calculate free energy profiles for microscopic rates and determine the microscopic binding pathway? While these issues are hard to elucidate experimentally, computational techniques offer the atomistic resolution needed.

Over time, multiple computational techniques have been developed to study ligand-protein interactions and ligand binding. Techniques such as docking (Taylor et al., 2002; Yuriev et al., 2011) aim at empirically determining binding sites and approximative affinities. Linear interaction energy (Hansson et al., 1998; Aqvist and Marelius, 2001; Åqvist et al., 2002) is an efficient semi-empirical technique to study binding energetics considering only intermolecular interactions between ligand and protein. More precise techniques try to estimate binding affinities using PMFs, either via implicit calculation (Cohen et al., 2006) or from MD simulations (Woo and Roux, 2005).

Especially recently MD simulations have also shown to be powerful tool to shed

light not only affinities but also on the pathways and kinetics of the binding process (Case and Karplus, 1979; Chang et al., 2007; Buch et al., 2011; Shan et al., 2011; Dror et al., 2011).

In this part we use MD simulations to investigate the kinetics and pathways of cAMP binding as well as the conformational changes in the CNBD of MloK1 to establish a complete picture of the binding process. For that we divided our investigations into two chapters:

In chapter 4 we will investigate the kinetics and the free energy landscape of the conformational change in the CNBD to classify the binding mechanism in the framework of conformational selection and induced fit. We will furthermore investigate the kinetics for the unbinding processes, which turns out to depend much on the kinetics of the conformational change, to derive an estimate for the effective off-rate.

In chapter 5, we will focus on the ligand's pathway to the binding site, i.e. how it exactly binds to the protein and how this relates to the experimentally determined on-rate.

4. The Conformational Change in the CNBD

4.1. Introduction

The conformational selection and induced fit mechanisms differ by the temporal order in which the conformational change and the ligand binding itself occur. To properly distinguish between the two concepts, we consider besides the open ligand-free conformation and closed ligand-bound conformation two additional potential binding substates, namely a closed ligand-free conformation and an open ligand-bound conformation (see figure 4.1).

Peuker et al. (2013) determined the on- and off-rate of cAMP at the wild type CNBD and two mutants to distinguish between conformational selection and induced fit. The assumption was that the mutations do not influence the kinetics of the ligand binding but the kinetics of the conformational change. As the mutations only affected the off-rate and not the on-rate, they concluded that the binding happens prior to the conformational change, i.e., that the binding is according to the induced fit mechanism.

Here we want to use a different approach: By directly calculating the free energy differences and barriers and subsequently the rates between the substates using MD simulations we distinguish between the two concepts. The substates are sketched in figure 4.1. Which model describes the ligand binding mechanism is governed mainly by the free energy differences between the open and closed conformation: In the conformational selection mechanism, both the closed conformation and the open conformation have to be energetically accessible in the ligand-free state and upon ligand binding the free energy minima shift towards the closed conformation (figure 4.1, bottom). In the induced fit mechanism, however, the closed conformation is not accessible prior to ligand binding and only becomes populated after ligand binding (figure 4.1, top).

To calculate the free energy landscape for the conformational change in the CNBD, a suitable reaction coordinate that describes this conformational change is required. The first goal is therefore to develop a method to derive a reaction coordinate for the conformational change. We then used this reaction coordinate to calculate the free energy differences and barriers between the open and the closed conformation of the CNBD in presence and in absence of the ligand to elucidate the binding mechanism by comparison with the prototypical energy landscapes sketched in figure 4.1.

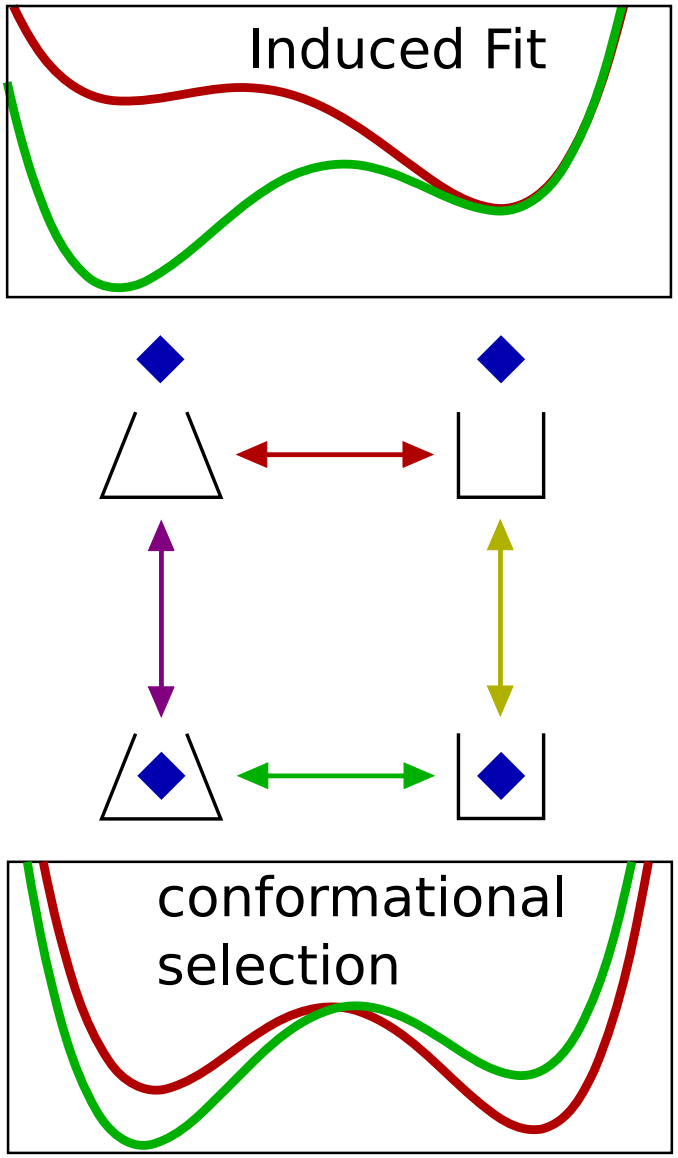


Figure 4.1.: Substates in a ligand binding process with conformational changes (middle) and shape of the free energy landscape in presence and absence of the ligand for an induced fit and a conformational selection binding mechanics.
The protein can adopt closed (left) and open (right) conformations and the ligand is either unbound (top) or unbound (bottom).
The induced fit and conformational selection model differ by the free energy landscape for the conformational change depending if the ligand is present (green) or absent (red): In the induced fit case (top), the closed conformation is not accessible prior to ligand binding. In the conformational selection picture (bottom) both conformations are accessible prior to ligand binding and the preference shifts upon ligand binding to the closed conformation.

Experimental measurements were so far unable to directly determine the microscopic rates for the conformational changes. Instead they were limited to the determination of a global on-rate and a global off-rate from the closed ligand-bound conformation. How the conformational change contributes to the off-rate, however, remained unclear.

To solve this puzzle and to simultaneously compare our estimates for the microscopic kinetics with experimental measurements, we determined the unbinding rates from the ligand-bound states and combined these with the rates of the conformational change to get an estimate of an effective global off-rate.

4.2. Theory

4.2.1. Derivation of a Reaction Coordinate for the Conformational Change

What are the requirements for a good reaction coordinate? Since we are interested in the conformational change in the CNBD, the different conformations should have separated values along the reaction coordinate. A second import requirement is that the transition state ensemble corresponds to a distinct value along the reaction coordinate that is different from the values of the two conformations.

We assume that there are two separated minima in the multidimensional free energy landscape that correspond to the two conformations. This is sketched in figure 4.2. If the vector connecting the centres of the probability densities corresponding to the two minima is used as a reaction coordinate, the resulting projections of the probability densities show a non negligible overlap. As a consequence, a PMF along this coordinate would show a barrier that is smaller than the actual free energy of the transition state. The best linear reaction coordinate, however, is obtained if projections of the two probability densities on the corresponding vector have the smallest overlap (figure 4.2 C). For the reaction coordinate defined by this vector the free energy of the barrier separating the minima is equal to the free energy of the transition state.

If the ensembles of the open and closed conformation are separated in the conformational space and if the shape of their probability distribution is at least very roughly given by multidimensional Gaussians, it is reasonable to assume that the optimal linear reaction coordinate is given by the vector for which the projections have the smallest overlap.

This vector is given as a the solution of an optimisation problem, where the quantity to minimise is

$$O(\mathbf{v}) = \int \mathrm{d}x \; \rho_a(x;\mathbf{v}) \cdot \rho_b(x;\mathbf{v}). \tag{4.1}$$

Here $\mathbf{v}$ is the vector onto which the probability densities are projected, ρ_a and ρ_b

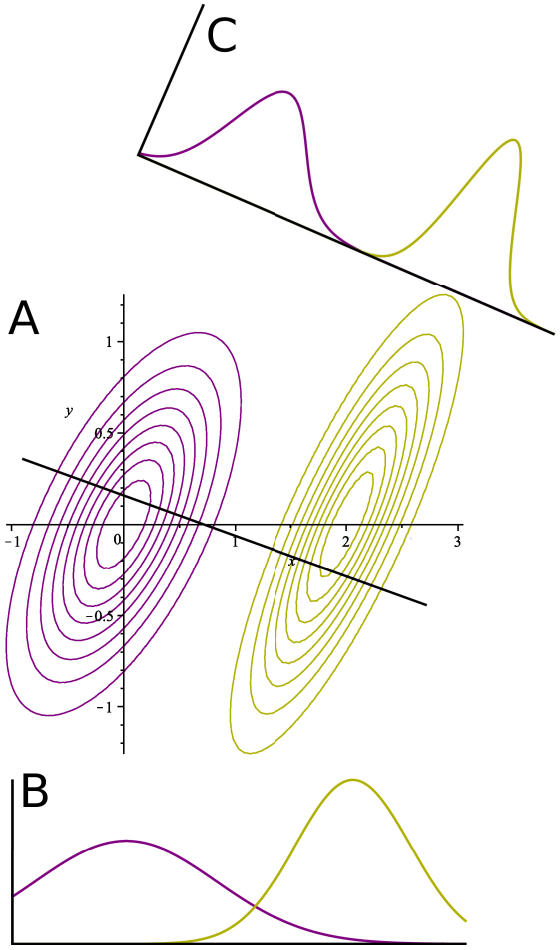


Figure 4.2.: A: Sketch of a two multi-dimensional probability densities B: Projections onto the vector connecting the two centres. C: Projections on the vector for which the maximal separation between the two densities is maximal.

are the projected densities:

$$\rho_a(x;\mathbf{v}) = \int \mathrm{d}^n q \ \rho_a(\mathbf{q}) \cdot \delta(\mathbf{v}\cdot\mathbf{q} - x). \tag{4.2}$$

As the probability densities are a priori unknown, they have to be obtained from MD simulations. For finite sampling the integrals consequently become sums over histograms. As the derivative of $O(\mathbf{v})$ with respect to the components of $\mathbf{v}$ cannot be obtained analytically, a gradient-free minimisation routine is employed to find the best vector $\mathbf{v}$. The implementation details are provided in section 4.3.2.1.

4.3. Methods

4.3.1. Molecular Dynamics Simulations

The simulations for the derivation of the reaction coordinate of the conformational change were set up and performed using Gromacs 4.0.7. (Van Der Spoel et al., 2005; Hess et al., 2008). The umbrella sampling (Torrie and Valleau, 1977) simulation for the calculation of the PMFs (Kirkwood, 1935) of the conformational change of the protein, the umbrella sampling simulations for the calculation of the PMF of the ligand unbinding and the free simulations starting at the open NMR structure were set up using Gromacs 4.0.7 and performed out using Gromacs 4.5 (Pronk et al., 2013). Unbiased simulations for the closed and open protein conformation where set up and performed using Gromacs 4.6.

In all simulations the Amber99sb (Hornak et al., 2006) force field was used together with the TIP3P water model. (Jorgensen et al., 1983). Electrostatic interactions were calculated using particle-mesh Ewald (Darden et al., 1993) with a real space cutoff of 1.2 nm, a grid spacing of 0.145 nm and cubic interpolation. Van-der-Waals interactions were cut-of at 1.2 nm. In simulations carried with Gromacs 4.6 the real space part of the electrostatic potential and the Lennard-Jones-Potential was shifted such that the potential was 0 at the cutoff point. For simulations performed with Gromacs 4.6, non-bonded interactions were calculated using Verlet neighbourlists (Verlet, 1967). The Verlet buffer was chosen such that the maximal error for pair interactions was less then $0.005\,{}^{\mathrm{kJ}}/_{\mathrm{mol}\cdot\mathrm{ps}}$. For all other simulations a group based cutoff scheme was employed van der Spool et al. (2005). All simulations were performed in the constant temperature, pressure and particle number ensemble, using the velocity rescaling method (Bussi et al., 2007) for temperature coupling with a heat path temperature of 300 K and a coupling time constant of 0.1 ps and Berendsen pressure coupling (Berendsen et al., 1984) with a reference pressure of 1000 hPa and a respective time constant of 1 ps. Simulations for determination of the reaction coordinate were carried out in cubic simulation boxes, all other simulations were carried out in dodecahedron shaped simulation boxes. All simulations were performed using periodic boundary conditions. All atom bonds were constraints using the LINCS algorithm (Hess et al., 1997), i.e. harmonic

pair bond interactions were removed from the force field potential and replaced by constraints. In simulations carried out with Gromacs 4.6 additionally degrees of freedom representing fast fluctuations of angles defined by at least one hydrogen atom were removed by changing the respective hydrogen atoms into virtual sites and fixing the angle (Feenstra et al., 1999). The equations of motion were integrated using the Verlet algorithm (Verlet, 1967) using a time step of 4 fs for simulations with virtual interaction sites and using a time step of 2 fs for all other simulations.

In all simulations carried out with Gromacs 4.6 sodium chloride ions where added at a concentration of 113 mmol/l. In all other simulations ions were added to ensure neutrality of the simulation box.

4.3.1.1. Ligand Parametrisation

Force field parameters for unprotonated cAMP were calculated using the general Amber force field (GAFF) (Wang et al., 2004) and the antechamber toolkit (Wang et al., 2006). Atomic point charges were obtained in a restrained fitting procedure in such a way that the resulting electrostatic potential fits best to the electrostatic potential generated by the electronic wave function (RESP charges) (Bayly et al., 1993). The wave function was calculated with Gaussian03 (Frisch et al.) at a Hartree Fock level using the 6-31G* basis set. A tight convergence criterion of 10^{-8} was applied for the self consistent field (SCF) calculations; 6 points per unit area were calculated in the electrostatic potential (ESP) fit. Before the calculation of the electronic wave function a geometry optimisation of the molecule was performed.

4.3.1.2. Modelling of the Protein Structure

All simulations starting in the open protein conformation were based on the R384A mutant (PDB ID code 1U12) (Clayton et al., 2004) of the CNBD. The only exception are the simulations started explicitly in the open conformation NMR structure (see section 4.3.1.5). The mutation was reverted by replacing residue Ala384 with Arg384 and missing atoms were added using the Modeller 9v10 software (Šali and Blundell, 1993). All simulations starting in the closed conformation were based on the crystal structure of the closed conformation (PDB ID code 1VP6). In both cases only one monomer was considered, and all ions were removed from the crystal structure.

4.3.1.3. Reaction Coordinate Determining Simulations

To derive a vector along which the projections of the open conformation ensemble and closed conformation ensemble separate best, preliminary sampling of open and closed conformation ensembles was needed. Therefore 4 50 ns unbiased simulations of the closed conformations were carried out, as well as 50 100 ns unbiased simulations of the open conformation.

4.3.1.4. Hamiltonian Replica Exchange Umbrella Sampling Simulations

For the construction of the PMF (Kirkwood, 1935) along the coordinate describing the conformational change Hamiltonian replica exchange umbrella simulations (Torrie and Valleau, 1977; Swendsen and Wang, 1986; Geyer, 1991; Fukunishi et al., 2002) were carried out (see sections 2.2.2 and 2.2.4).

25 umbrella windows were simulated for 250 ns each. This was done both for systems with and without bound ligand. In each umbrella window a biasing potential $U_i(x) = \frac{\alpha}{2}(x - x_{\text{ref},i})^2$ was applied, where x is the projection of the current conformation onto the vector serving as the reaction coordinate, $x_{\text{ref},i}$ the reference point for the corresponding umbrella potential and $\alpha = 10\,{}^{\text{kJ}}/_{\text{mol}\cdot\text{nm}^2}$ the spring constant of the umbrella potential. The reference points were spaced equidistantly along the reaction coordinate from -6.5 nm to 5.5 nm. Every 100 ps replica exchange attempts between 10 randomly chosen pairs of umbrella windows were performed.

The starting conformations where constructed as follows:

For equilibration, starting both from the closed conformation as well as from the open conformation, systems with a restraining potential identical to the one used in the umbrella sampling simulations were simulated for 20 ns. The end structures were then used as starting conformations for simulations with restraining potentials corresponding to the umbrella potentials of neighbouring umbrella windows. Following this procedure, the systems was "dragged" from the x-ray conformations along the reaction coordinate.

The actual starting conformation for the umbrella sampling simulations were then taken from the end conformation of these preequilibration simulations. The starting structures of the 8 umbrella windows closer to the closed conformation were taken from the preequilibration run starting at the closed conformation, while the starting conformation for the remaining umbrella windows were extracted from the pre-equilibration simulations starting at the open conformation.

4.3.1.5. Simulations Starting in the Open Conformation NMR Structure

To test if the open conformation ensemble is biased by the choice of the starting structre, 9 50 ns simulations were started in the open conformation NMR structure (PDB code 2kxl).

4.3.1.6. Unbiased Simulations

Simulations to estimate rates by counting transitions were carried out for the open conformation and closed conformation both in presence and absence of the ligand.

180 1 µs simulations were carried out starting in the closed ligand-bound conformation, using the ligand-bound x-ray structure as starting structure.

20 200 ns simulations were carried out starting in the closed but ligand-free conformation, using the same starting structures after removal of the ligand.

40 1 µs simulations were carried out starting in the ligand-free open conformation of the x-ray structure.

99 1 µs simulations were carried out starting in the open conformation with a ligand molecule placed inside the binding pockets using coordinates extracted from the ligand-bound x-ray structure.

4.3.1.7. Umbrella Sampling Simulations of Ligand Unbinding

To calculate the PMF for the unbinding of the ligand from the closed conformation, umbrella sampling simulations were performed using the distance between the ligand centre of mass (COM) and the COM of the binding site surrounding residues GLY297, ARG307 and SER308 as a reaction coordinate. 30 umbrella windows were placed equidistantly along the reaction with a distance of 0.05 nm. A spring constant of 3000 $^{kJ}/_{mol \cdot nm^2}$ was used.

The starting configurations were derived sequentially by simulating the previous umbrella window for 10 ns and using the final configuration as the starting configuration of the next umbrella window.

The umbrella windows where the ligand was still in the binding site were simulated for 50 ns, intermediate windows for 100 ns and all others for 200 ns.

The PMF was calculated using the WHAM (Kumar et al., 1992) (see section 2.2.3).

4.3.2. Data Analysis

4.3.2.1. Derivation of a Coordinate Describing the Conformational Change

The vector that describes the conformational change was obtained by minimising equation 4.1.

To accelerate the search, the search for the $3N$ dimensional vector was not performed in the full configurational space but limited to a much smaller subspace. To that end a PCA (see section 2.3) on the free simulations of the open and closed conformation (described in section 4.3.1.3) was performed. Subsequently the projections onto the three eigenvectors corresponding to the three largest eigenvalues were calculated.

The minimisation of equation 4.1 was carried out using the downhill simplex method by Nelder and Mead (1965). Since the vector defining the reaction coordinate is supposed to be normalised, in every iteration of the simplex method the newly obtained vector was normalised.

To evaluate if the restriction of the search space to the space spanned by the first three eigenvectors of a PCA is appropriate, the same procedure was repeated searching in the subspace spanned by the first four, five, six, seven and eight eigenvectors. For the obtained vectors the dot product with the optimal vector in three PCA-dimensions is calculated.

4.3.2.2. Calculation of Potentials of Mean Forces of Conformational Change

The one-dimensional PMF along the reaction coordinate was calculated with the WHAM. The weight of every conformation, consisting of the value of the biasing potential for the given conformation and the free energy value associated with the umbrella window (see section 2.2.2) was extracted. Subsequently every conformation occurring in the umbrella sampling simulations was projected not only on the reaction coordinate but also on the coordinate orthogonal to the reaction coordinate where the largest variance was observed. Using the previously obtained weights, for these two coordinate a weighted histogram was calculated and transformed into an two-dimensional PMF. This procedure was applied both to the system with and without bound ligand.

4.3.2.3. Rate Estimation

To avoid counting of spurious transitions events, a conformational change in an unbiased simulation was considered as such if the system progressed along the reaction coordinate to the value for which the minimum of the other conformation is found in the PMF and stayed there for the remainder of the simulation.

Unbinding events were counted as such, if the COM distance between the ligand and the binding site surrounding residues GLY297, ARG307 and SER308 was larger than 1.2 nm. As the equilibrium distance in the bound conformation is 0.6 nm, this corresponds to displacement of 0.6 nm from the bound state.

For both cases rates were then calculated as the quotient of totally observed transitions n and accumulated dwell time in the starting state t: $k = \frac{n}{t}$. Statistical errors were estimated via $\sigma_k = \frac{\sqrt{n}}{t}$. The discrepancy of the symmetric standard deviation error estimate from the asymmetric errors defined by a 68,3% confidence interval was found to be smaller than 1.4 even in the most extreme case and therefore not considered.

See section 2.4 for a derivation of the theoretical background.

4.4. Results and Discussion

4.4.1. Derivation of a Reaction Coordinate

We aim at categorising the binding process in terms of induced fit and conformational selection by determining the free energy differences and barriers between the open and closed conformation of the CNBD. For this goal it is necessary to first construct a reaction coordinate that describes the conformational change (see sections 4.2.1 and 4.3.1.3).

The optimised vector was determined in the three-dimensional spaced spanned by the three eigenvectors corresponding to the largest eigenvalues of a PCA on unbiased simulations. Figure 4.3 shows the projection of the simulation data used

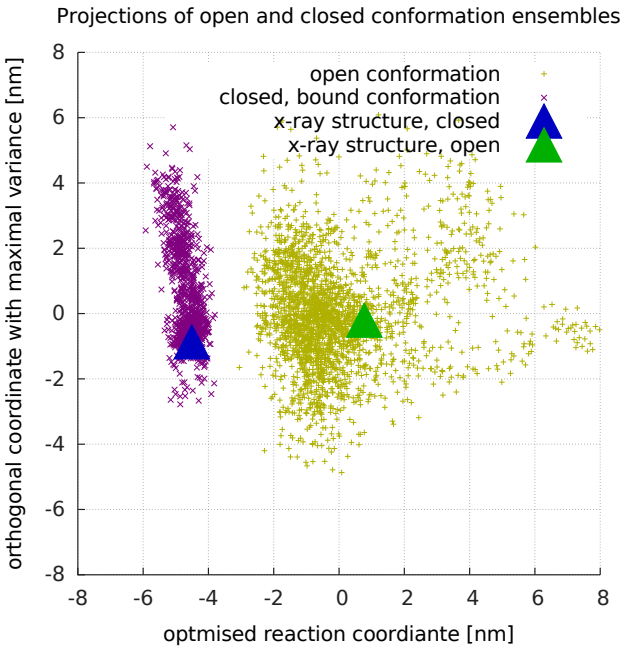


Figure 4.3.: Projections of the closed conformation ensemble (violet) and open conformation ensemble (yellow-green) used to determine the optimised reaction coordinate projected onto that coordinate and the orthogonal coordinate for which the largest variance is observed.

to determine the reaction coordinate onto the reaction coordinate and the coordinate orthogonal to it where the largest variance within the ensembles occurs. The datasets are well separated, rendering the coordinate suitable to describe the conformational change.

To cross-validate the restriction to three dimensions the vector yielding optimal separation between open and closed conformation ensemble was determined within a higher dimensional subspace spanned by more PCA vectors. The dot product of these vectors with the one determined in three dimensions was however always larger than 0.9. Given the high dimensionality of the full $3N$ dimensional space, these vectors are thus considered as highly similar.

4.4.2. Free Energy Landscapes

Next we addressed the question whether the binding process is best described by an induced fit or conformational selection mechanism by determining the free energy landscape of the conformational change.

To that end we performed umbrella sampling simulations of the CNBD in presence and absence of cAMP and calculated the PMF along the optimised reaction coordinate describing the conformational change in the CNBD both in absence and in presence of the ligand in the binding site (figure 4.4).

The PMF in presence of the ligand shows a clear minimum for the closed conformation and a second, smaller minimum for the open conformation. Defining the border between open and closed state as the maximum of the free energy barrier between them the free energy difference is $\Delta G = 8.9 \pm 0.3\,{}^{\mathrm{kJ}}/_{\mathrm{mol}}$.

The minimum for the closed conformation furthermore coincides with the ligand-bound x-ray and NMR structures. This is in good agreement with the expectation that the closed conformation dominates in presence of the ligand.

The PMF for the system in absence of the ligand shows one extended minimum for the open conformation. There is no distinct second minimum for the closed conformation, and applying the same cutoff between open and closed conformation yields a free energy difference of $\Delta G = -12.9 \pm 0.2\,{}^{\mathrm{kJ}}/_{\mathrm{mol}}$.

As expected the open conformation is preferred in absence of the ligand. The large free energy difference between open and closed conformation shows that the closed conformation will be barely populated in absence of the ligand. This strongly suggests that binding does not occur via a conformational selection, but via an induced fit mechanism. Also the shape of the resulting PMFs resembles very much the landscape as we would expect it for an induced fit mechanism (figure 4.1).

Whereas the minimum for the closed conformation is very narrow and coincides well with x-ray and NMR structures, the minimum for open conformation in the ligand-free system is much wider. Interestingly, the minimum is close to the x-ray structure of the R384A mutant, whereas the NMR structures are slightly more off.

Simulations starting with the open NMR conformation also show progress along the reaction coordinate towards the minimum observed in the PMF. This suggests

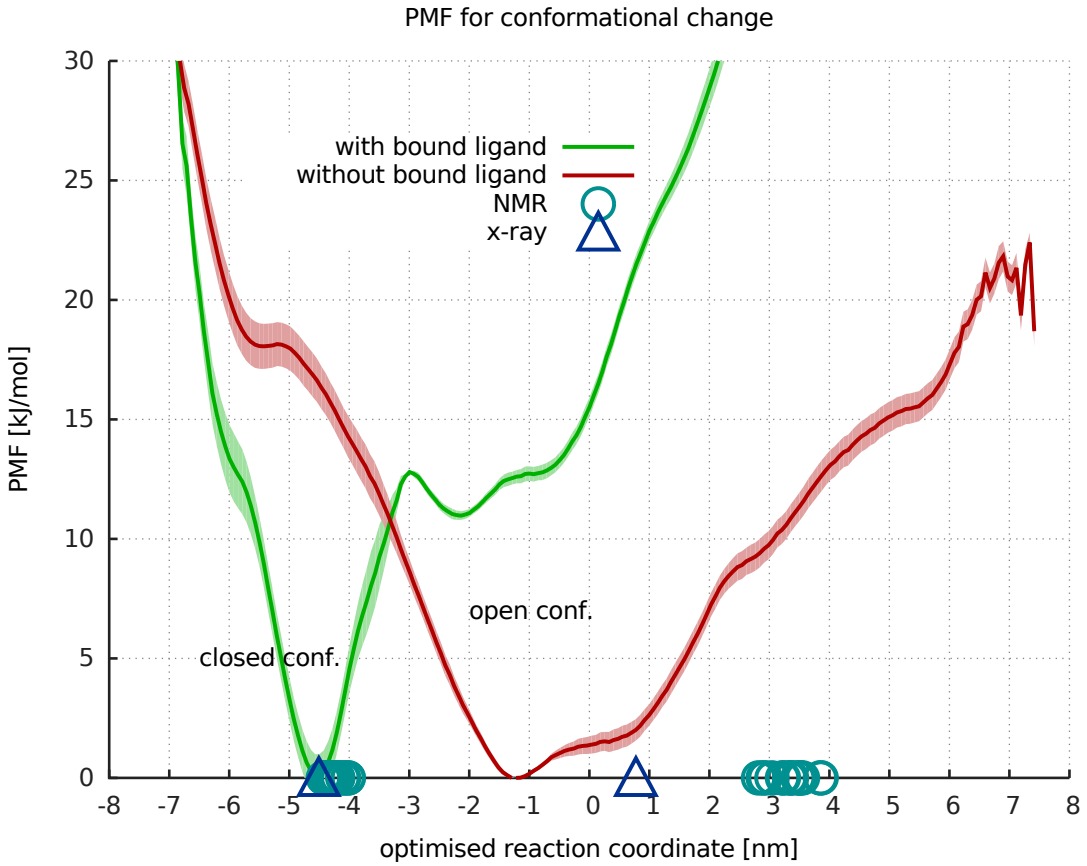


Figure 4.4.: PMF for the conformational change along the optimised reaction coordinate in presence (green) and absence (red) of the ligand.

that the NMR conformations do indeed not constitute a minimum during the MD simulations.

There is no obvious explanation for the discrepancy. Instead several possibilities exist: The difference could be caused by limited accuracy of the force field used in the MD simulations. This however would not suffice to explain the difference between the NMR and x-ray structure that was previously attributed to the crystallisation conditions (Schünke et al., 2011). Keeping in mind the good agreement of the minimum in the PMF and the x-ray conformation it would be a weird coincidence if crystallisation conditions and force field inaccuracies lead to the same offset. Thus it could also be that the conditions in the NMR experiment, e.g. the presence of non sodium chloride ion types, influence the NMR structure. There is also a third option: As the simulations already revealed that the open conformation exhibits large flexibility, it is possible that the different structures are the result of biases in the averaging during structure reconstruction. In any case the discrepancy is most likely not of extreme concern, as due to the large flexibility of the open conformations the NMR structure actually also belongs to the ensemble of sampled configurations in unbiased MD simulations of the open conformation.

What are the free energy barriers between open and closed conformation? The one-dimensional PMF for the ligand-bound system in figure 4.4 suggests that the barrier for an opening is roughly $\Delta G^{\dagger} = 13\,{}^{\mathrm{kJ}}/_{\mathrm{mol}}$, an the back barrier $\Delta G^{\dagger} \approx 2\,{}^{\mathrm{kJ}}/_{\mathrm{mol}}$.

However, it is unclear if the maximum in the one-dimensional PMF corresponds really solely to the transition state. If it does, the barrier corresponds to the real barrier between open and closed ensemble. Conversely if that is not the case, the apparent maximum in the PMF between the two states does not constitute the real free energy barrier but underestimates the true free energy of the transition state.

To resolve this issue, we did a first step by adding another dimension of the configurational space and calculated a two-dimensional PMF for the ligand-bound system from the umbrella sampling simulations using the reaction coordinate and the orthogonal coordinate with the highest variance (figure 4.5).

Both closed and open conformation are well recognisable as separate states. However it becomes apparent that the reaction coordinate does not provide a perfect separation between the two states, despite the fact that the reaction coordinate defining vector was derived such that a maximal separation between the sampled open and closed conformation ensembles was reached (see sections 4.2.1, 4.3.2.1 and 4.4.1). This derivation however was carried out using limited sampling of open and closed conformation ensembles, leading to a vector that does not perfectly separate the true "infinite sampling" ensembles.

In the two-dimensional PMF the apparent barrier between closed and open state in the two-dimensional PMF is $\Delta G^{\dagger} = 22\,{}^{\mathrm{kJ}}/_{\mathrm{mol}}$.

Obviously it could very well be that this estimate for the free energy barrier is also an underestimation of the true barrier since in the two-dimensional PMF all remaining orthogonal degrees of freedom are also averaged out. However, due to the increased statistical error accompanying higher dimensional histograms and

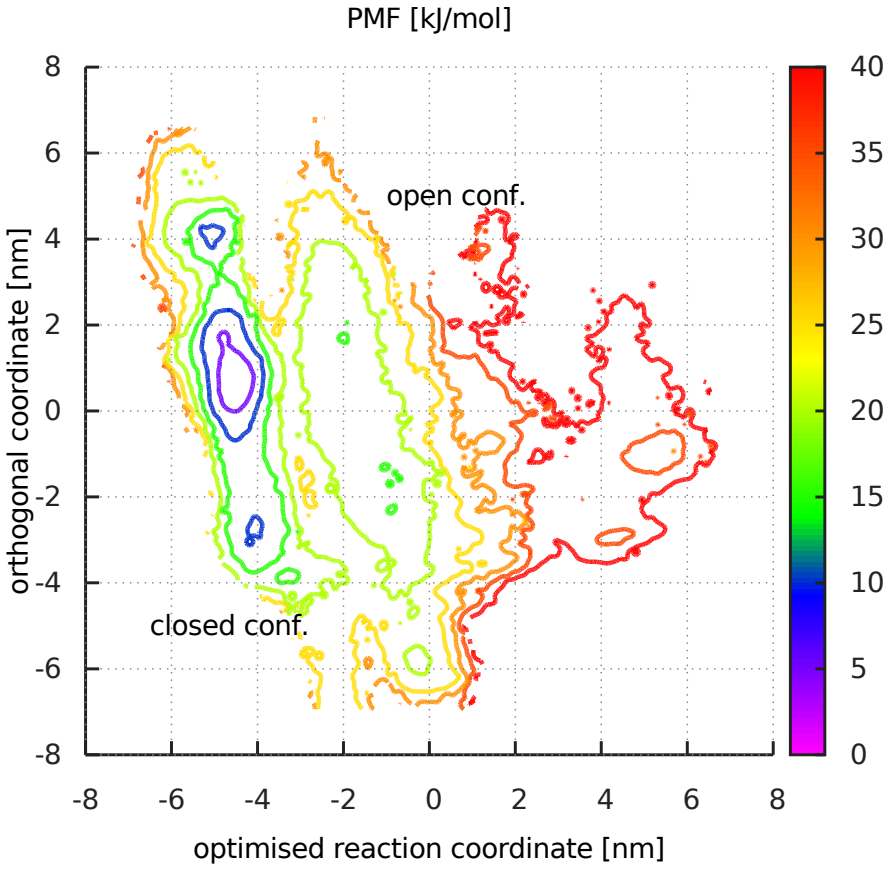


Figure 4.5.: Two-dimensional PMF of the conformational change for the ligand-bound system. The x-axis is the optimised reaction coordinate for the conformational change, the y-axis is the orthogonal coordinate where the highest variance is observed.

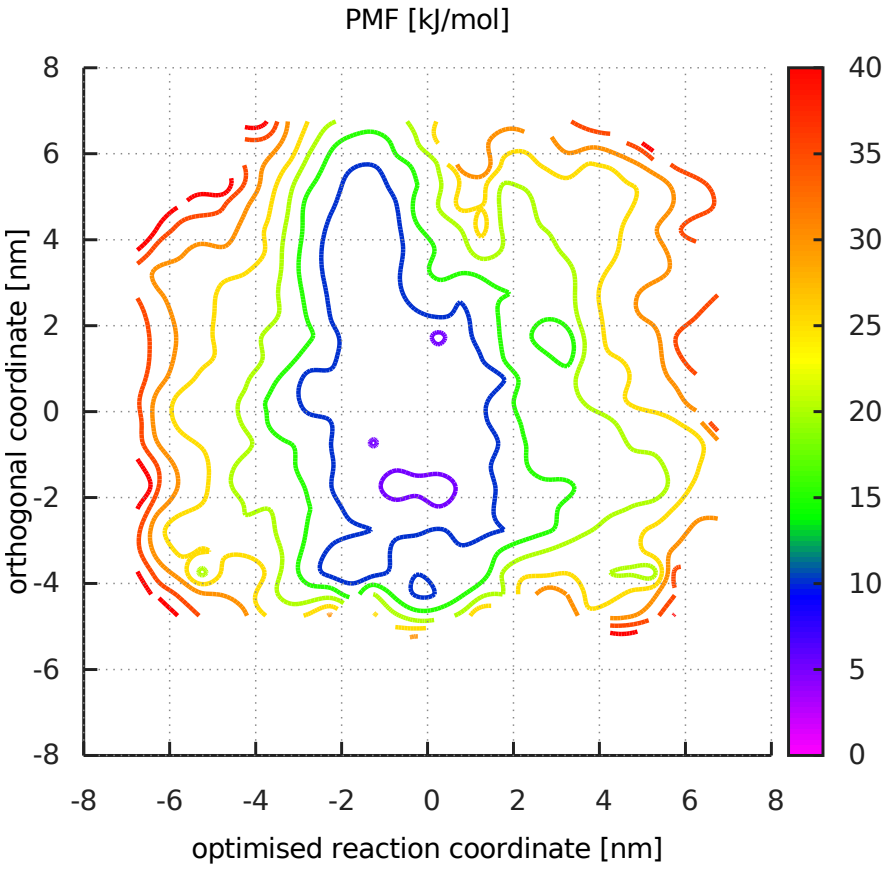


Figure 4.6.: Two-dimensional PMF of the conformational change for the ligand-free system.

the subsequent impossibility to correctly assign isosurfaces it is not feasible to extend this approach to three- or even more dimensional PMFs. Instead the barrier estimate should be interpreted as a lower boundary for the real free energy barrier.

The free energy minima of open and closed conformation in the two-dimensional PMF have similar extent in the second coordinate. In the direction of the reaction coordinate however the closed conformation is much more restrained. Apparently the presence of the ligand therefore does not constrain all motions, but mainly the opening-closing motion.

The two-dimensional PMF along the same coordinates for the ligand-free CNBD is depicted in figure 4.6. While the accessible configurational space volume is apparently much larger than for the ligand-bound system in the closed conformation, meaning that the protein exhibits more flexibility, no relevant free energy barriers orthogonal to the reaction coordinate are determined.

Therefore the open conformation is best interpreted as one extended minimum without relevant substates.

4.4.2.1. Cross-Validation of the Umbrella Sampling Simulations

How accurate are the calculation of the PMFs of the conformational landscape of the umbrella sampling simulations? To cross-validate our findings we extensively simulated the CNBD with and without ligand in the closed and open conformation, respectively.

From these equilibrium simulations a PMF on the same coordinates as of the umbrella sampling simulations is calculated (figure 4.7a and 4.7b). The comparison reveals that the equilibrium ensemble samples the region of the minimum of the PMF both for the system with and without bound ligand. Especially the agreement of the equilibrium ensemble of the open ligand-free conformation with the PMF is striking (figure 4.7a), boosting the confidence of the PMF of the ligand-free system.

The free energy minimum of the closed ligand-bound system also agrees roughly with the umbrella sampling PMF. However, the unbiased simulations appear to sample also parts of the open conformation though. As the system always quickly returns to the closed conformation minimum and no permanent opening is observed in the unbiased simulations, these outliers do not constitute real transitions though. Instead this finding indicates again that the reaction coordinate does not perfectly separate the two ensembles and only the projection onto the two coordinates creates the impression of "ghost transitions".

4.4.3. Rates of Conformational Changes

Next we addressed the question how fast transitions between open and closed conformation occur in presence and absence of the ligand. This was done both using the PMFs as well as counting transitions from unbiased simulations of the individual conformations.

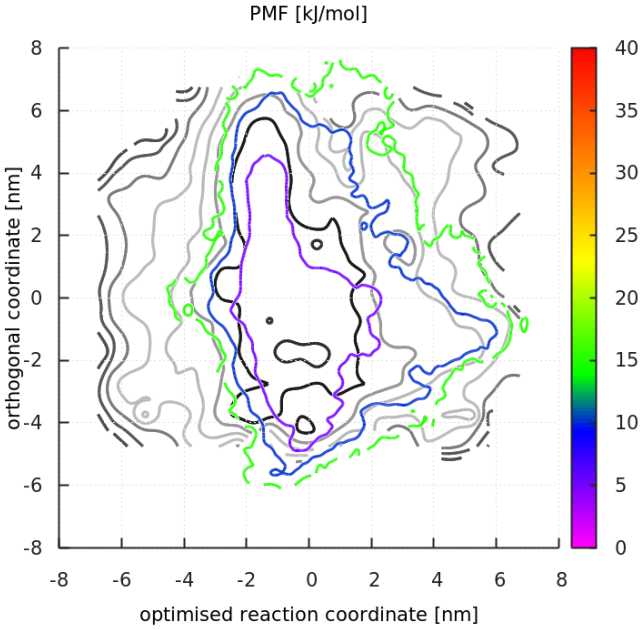
To obtain transition rate estimates from transition counting we performed simulations of the four substates, i.e. open and closed conformation with and without bound ligand. To avoid overestimation, transitions are only counted if the system stays in the other conformation for the reminder of the trajectory.

To obtain rates from the PMFs, we used the relation

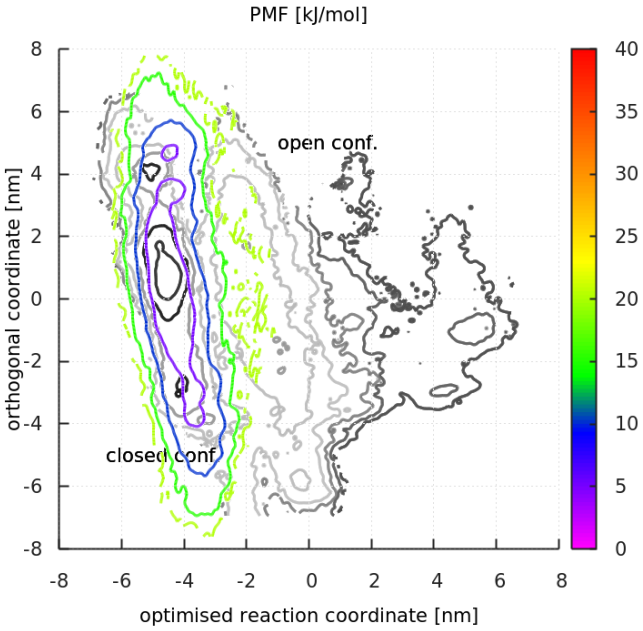
$$k = \omega \cdot \exp(-\beta \Delta G^{\dagger}) \tag{4.3}$$

with the attempt frequency ω. To obtain an estimate for the attempt frequency for all unbiased trajectories, the autocorrelation time for the projection onto the reaction coordinate was calculated and averaged. The resulting attempt frequency was $\omega := \tau_{ac}^{-1} = 25 \cdot 10^6\,\mathrm{s}^{-1}$, which is in the range of common attempt frequency for conformational changes in protein.

The resulting rate estimates are tabulated in figure 4.8.



(a) ligand-free open conformation



(b) ligand-bound closed conformation

Figure 4.7.: Coloured: Two-dimensional PMF of the conformational change from equilibrium simulations of the ligand-free open conformation (a) and ligand-bound closed conformation (b). The contour lines of the PMF derived from the umbrella sampling simulations without (a) and with (b) ligand are depicted in grey.

For the ligand-free systems the rates from transition counting and from the PMF qualitatively agree. Though it is difficult to estimate this rate from the PMF because the free energy landscape suggests that the process is a downhill process; interpreting the attempt frequency as a upper limit leads to an estimate of $k_{\text{open}} < 25000\,\text{ms}^{-1}$. Transition counting helped to render this estimate more precisely with a rate estimate of $k_{\text{open}} = (3100 \pm 1500)\,\text{ms}^{-1}$. Together with a closing rate estimate of $k_{\text{close}} \approx 60\,\text{ms}^{-1}$ this means that the closed ligand-free conformation is more than a factor of 100 less likely than the open ligand-free conformation. Together with the computational observation of spontaneous ligand binding at the open conformation (see chapter 5) this effectively rules out a conformational selection mechanism.

For the ligand-bound system, transition counting yielded a rate estimate of $k_{\text{close}} = 60 \pm 30\,\text{ms}^{-1}$, which is consistent with the upper limit obtained from the PMF of $k_{\text{close}} < 200\,\text{ms}^{-1}$.

As no opening from the ligand-bound closed conformation to the ligand-bound open conformation was observed in the unbiased simulations, only an upper limit of $k_{\text{open}} < 5\,\text{ms}^{-1}$ could be obtained (see section 2.4). Since the PMF also only yielded an upper limit of $k_{\text{open}} < 2.5\,\text{ms}^{-1}$, the exact rate remains (within the limits set by this estimate) uncertain.

4.4.4. Rates of Unbinding

What is the effective off-rate and how does the conformational change influence it? To that end we investigated how fast the ligand unbinds from the open and closed conformation.

The rate for unbinding from the open conformation (represented by the yellow arrow in figures 4.1 and 4.8) was calculated by counting transition into the unbound state, very similar to the rate estimation of the conformational changes. We thereby obtained a value off $k_{unbind} = 53\,\text{ms}^{-1}$.

To address the unbinding from the closed conformation (represented by the violet arrow in figures 4.1 4.8) we calculated the PMF along the distance of ligand to the binding site starting from the closed conformation (figure 4.9). The PMF shows a very high barrier for unbinding of $\Delta G^{\dagger} = 90\,{}^{\text{kJ}}/_{\text{mol}}$. With an attempt frequency of $\omega = 1\,\text{ns}^{-1}$ this barrier corresponds to a extremely slow rate of $k_{\text{unbind}} = 2 \cdot 10^{-7}\,\text{s}^{-1}$.

The drastic difference in the unbinding rates as of consequence of different protein conformations is most likely due to steric clashes as the ligand is pulled out of the binding pocket while the protein occupies the closed conformation.

Since unbinding form the closed conformation is negligible, the effective off-rate is determined by the rates for the conformational change in presence of the ligand and the unbinding rate from the open conformation. Combining the upper limit of the rate estimates for the conformational change in presence of the ligand and the unbinding rate estimate from the open conformation yields an estimate for the "effective" off-rate: $k_{\text{off,effective}} < 0.5\,\text{ms}^{-1}$.

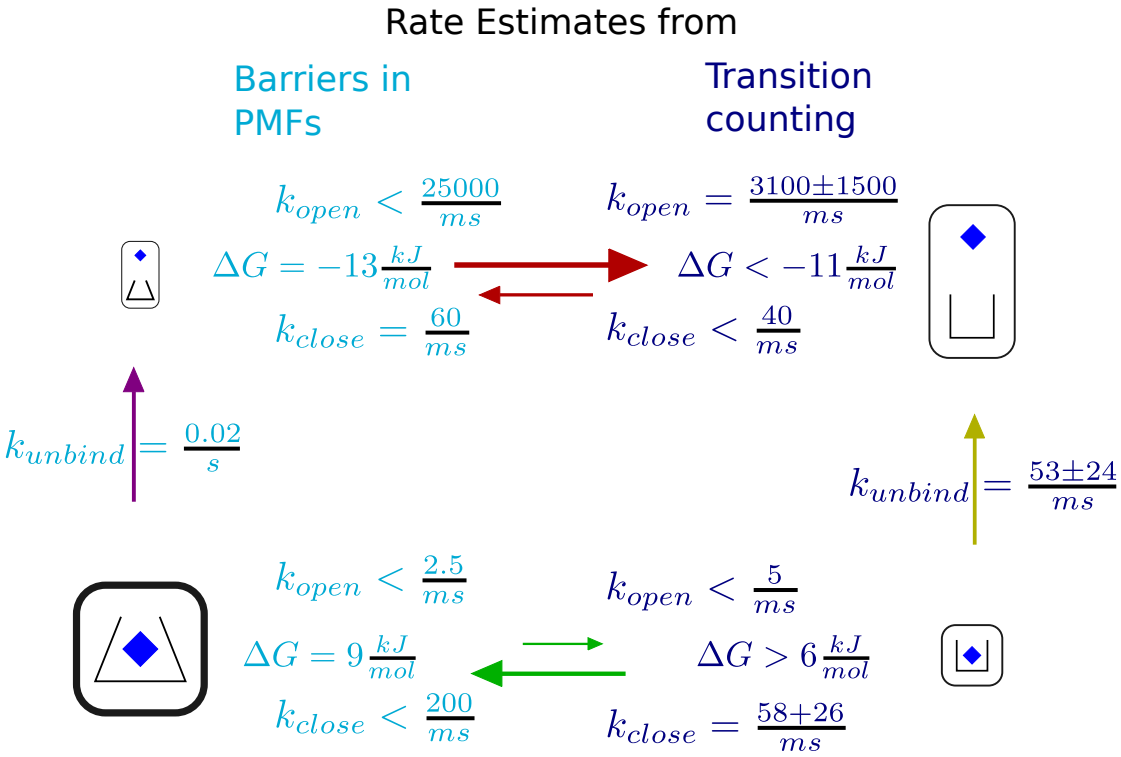


Figure 4.8.: Estimates for the rates and free energy differences of the conformational changes and unbinding. The rates and free energy differences between the substates were either calculated via the barriers in the PMFs (cyan, left) or by counting transition between the substates in unbiased simulations (blue, right).
The rate for unbinding from the closed conformation was also calculated via a PMF (left), whereas the rate for unbinding from the open conformation was determined via transition counting (right).

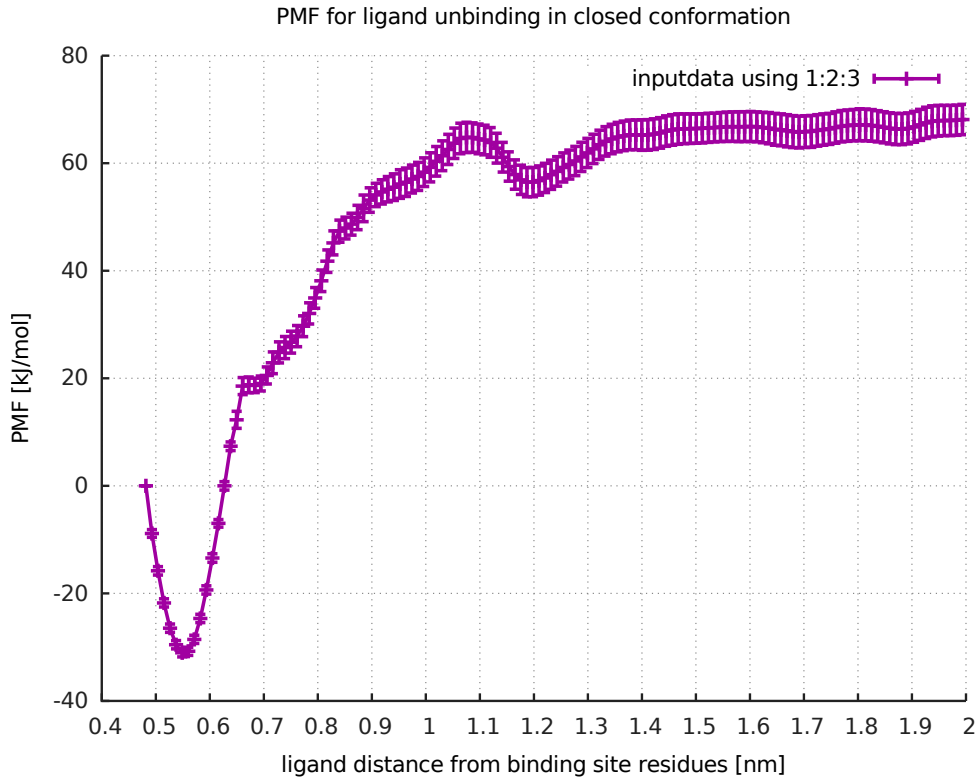


Figure 4.9.: PMF for ligand unbinding in the closed conformation. The distance of the COM of the ligand from a group of binding site surrounding residues is used as a reaction coordinate.

Apparently the effective off-rate is dominated by the slow opening rate of the ligand-bound protein. The upper limit for the effective off-rate is much faster than the experimentally determined value of $k_{\text{off, exp}} = 1.76\,\text{s}^{-1}$ (Peuker et al., 2013). Although our estimate for the upper limit is thus only a weak prediction, it is still consistent with the experimental value.

If the pathway from the closed ligand-bound conformation to the open ligand-free conformation via the closed ligand-free conformation has a negligible weight, whereas the pathway via the open ligand-free conformation is relatively easy accessible, then, due to microscopic reversibility, the same must hold for the binding pathway. This means that binding has to occur before the conformational change, which is again a confirmation of the induced fit interpretation.

4.5. Conclusion

In this chapter we studied the kinetics of the conformational change in the CNBD of MloK1 during the binding of cAMP to classify the binding mechanism either according to the induced fit or according to the conformational selection picture, predicted a set of microscopic internal rates determined an upper limit for the effective off-rate.

Motivated by the available protein structures, we first postulated a simple 4-state model consisting of ligand-bound and ligand-free states as well as open and closed receptor conformations. We then developed a method to determine a linear coordinate describing the transition between the receptor conformations that aims at separating free energy minima and the transition state ensemble along a one-dimensional coordinate.

The PMFs along this reaction coordinate revealed that the closed conformation is indeed a free energy minimum if the ligand is bound. They furthermore revealed that, in absence of the ligand, the closed conformation in equilibrium conditions is barely populated, strongly suggesting an induced fit mechanism. Additionally, it was revealed that the open conformation is more flexible than the closed conformation. It is very likely that interactions formed in the presence of the ligand reduce this flexibility.

We cross-validated our findings on the free energy landscapes by calculating the rates between the four postulated substates by counting transitions occurring in unbiased simulations of the substates. The rates estimated from this transitions counting agree up to an order of magnitude with the rates derived from the PMFs, justifying our approach and increasing the confidence in the obtained microscopic rates and in the interpretation of the binding mechanism as an induced fit mechanism.

This classification of binding mechanism as an induced fit is also in perfect agreement with experimental predictions. (Peuker et al., 2013).

The opening of the closed ligand-bound conformation was found to be the slowest transition with a rate whose upper border is in the millisecond regime. The closing of the open conformation was estimated to happen on a multi-microsecond timescale. As the measurement of this transition until now escaped experimental investigation, our estimate constitutes a novel prediction. The opening of the closed conformation in absence of a bound ligand happens very rapidly on a sub-microsecond timescale, underscoring the instability and negligible role of this conformation.

To investigate which unbinding pathway contributes to the effective off-rate that was measured in experiments, we calculated the unbinding rates from the ligand-bound closed conformation and the ligand-bound open conformation. The rate estimate for direct unbinding from the closed conformation was $0.02\,\mathrm{s}^{-1}$, whereas the unbinding from the open conformation was estimated to be $50 \pm 25\,\mathrm{ms}^{-1}$. These findings strongly suggest that binding or unbinding in the closed conformation does not play a relevant role, further supporting the induced fit picture.

Combining the rates of the conformational change with the unbinding rates we obtained an upper limit for the effective off-rate of $0.5\,\mathrm{ms}^{-1}$.

This upper limit is consistent with the experimentally measured off-rate, however it is also by a factor of 250 larger than the experimental result. This indicates that the rate for the conformational change from the ligand-bound closed conformation to the open conformation actually happens with a much lower rate than the upper limit we obtained. Further refinement of the reaction coordinate, as well as more extensive sampling might help to lower and refine this estimate.

Our approach was to decompose the binding process into different substates according to protein conformation and ligand presence/absence, to systematically derive a linear reaction coordinate for the conformational change, to calculate the free energy profile for this coordinate in presence and absence of the ligand and to cross-validate these results with long unbiased simulations.

It would be interesting to find out in future work if and how well this generic approach is transferable to other systems where ligand binding is connected with conformational changes. If it turned out that the approach is applicable to other systems as well, a refined and automated version of the techniques sketched here might be applied systematically to a large group of proteins to predict a binding mechanism classification and, using the results and techniques presented in the next chapter, a pair of effective on- and off-rates.

5. The Binding of cAMP

5.1. Introduction

In the previous chapter, the interplay of ligand binding and conformational change was investigated, with focus on the timescales of the conformational changes. However, the exact pathway of the ligand to the binding site as well as the effective on-rate were not discussed. These topics will be investigated in this chapter.

To measure the on-rate of the ligand binding process, Peuker et al. (2013) have performed stop-flow experiments with the CNBD of MloK1 and a dye labelled analogue of cAMP, 8-NBD-cAMP. The fluorescence of 8-NBD-cAMP is quenched in aqueous solvents and high in a hydrophobic environment. After mixing protein and ligand, the increase in fluorescent intensity was recorded over time. Assuming that fluorescence is high in the binding pocket and low everywhere else, the dependence of the fluorescent intensity on time was then used to derive the on-rate of the binding process.

In a simple Smoluchowski model (Smoluchowski, 1917) (see also section 2.5) the on-rate constant is determined by radial diffusion to the protein surface. Applied to the CNBD of MloK1, such a model would yield the same result as the experiment, if either the labelled ligand became fluorescent already at the protein surface or if a ligand that reaches the protein surface progressed in a much shorter time to the binding site than the surface binding had taken.

However, the measured on-rate constant of $k_{\mathrm{on}} = 26 \frac{1}{\mu\mathrm{s}\cdot\mathrm{mol/l}}$ is more than two orders of magnitude smaller than the simple Smoluchowski approximation.

An obvious explanation for this discrepancy would be that the assumptions made above are not fulfilled: Instead, the ligand is fluorescent only in the binding pocket itself and the ligand has to diffuse directly into a the binding site. This explanation would mean that the binding is described by Smoluchowski diffusion not to the protein surface but to a smaller volume representing the binding site. However, if this were the case, the target volume according to the Smoluchowski model would have to have a radius of approximately 10 pm to be compatible with the measured on-rate constant. Such a small target radius is very difficult to imagine, suggesting that the actual binding process is more complex.

In this chapter we want elude the detailed dynamics of the ligand to explain the large discrepancy between the naive Smoluchowski approximation and the actually observed on-rate.

One possible explanation is that it is not only radial diffusion that determines the on-rate. Instead it is possible that the binding is governed by a binding funnel

which constrains both translation and rotation. If the ligand forms attractive interactions at the bottom of the funnel with the protein, the sum of enthalpic and entropic contributions from attractive interactions and confinement lead to a free energy barrier in the funnel. Such binding funnels, similar to funnels observed in protein folding, have already been postulated and observed for ligand binding (Wang and Verkhivker, 2003; Shan et al., 2011).

It is, however, in principle also possible that the essential binding pathways *are* via the protein surface, i.e., that the ligand binds to the surface or at least to certain surface spots and then rolls via other surface states into the binding site. In this case, if the binding to surface is diffusion limited, it would have to be the surface rolling and detachment rates that lead to the low overall on-rate.

To decide how the ligand reaches the binding site and what determines the on-rate constant, we first constructed a unified model of the binding process that contains both of the aforementioned features, namely a binding funnel as well as surface states and pathways. For this model (which is described in detail in section 5.2.1) we assumed that the surface and especially the space separating bulk phase and binding site is partitioned into an unknown number of states and that the dynamics of the ligand is given by transitions between these states. In this unified model binding via funnel or binding via surface rolling are simply limiting cases for different sets of statistical weights of the states and transition probabilities between the states. It then depends on these weights and transition probabilities which binding pathway is dominating.

To find out which model explains the measured on-rate we carried out 900 200 ns simulations of cAMP and the ligand free CNBD in solution in which spontaneous binding occurred. We then identified the relevant surface states by constructing a fine-gridded Markov model of the ligand dynamics. We then went on to test if transitions from these surface states into the binding site constitute a relevant binding pathway. After ruling out this pathway, we determined how well the funnel model describes the ligand binding process and then used the funnel model to predict an on-rate constant.

5.2. Theory

5.2.1. Substates and Pathways in a Ligand Binding Process

In a traditional picture, ligand binding is described by a simple two-state process with one bound state and one unbound bulk-phase state. This model has yielded very satisfactory results for low ligand concentrations, however they often, e.g. in this case of cAMP binding at MloK1, do not coincide with on-rates predicted by a simple Smoluchowski model where only diffusion to the surface is considered.

Microscopically however, there are many substates, e.g. surface bound states, multiple ligand orientations or protein conformations. Such substates and potential transitions between them are sketched in figure 5.1. In such a multi-state system,

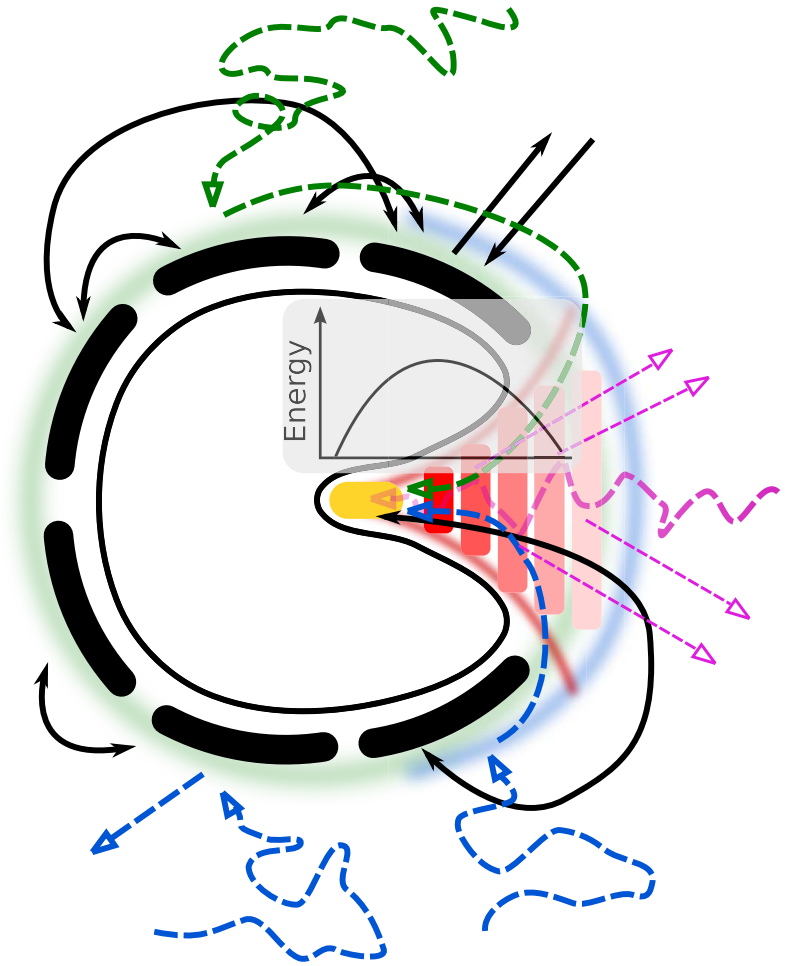


Figure 5.1.: Potential substates, transitions and pathways for ligand binding. The ligands conformational space is separated into bulk phase states, a binding site states (yellow, rounded) and surface states (black, rounded) with associated transition probabilities (black arrows). There are multiple hypothesis for the ligand binding pathway: Either a two-step process to the protein surface (green blur) or part of the surface (blue blur), followed by surface rolling into the binding site (green and blue arrows, respectively), or via entering a binding funnel (red blur), followed by a stochastic barrier crossing (pink arrows))

the dominant ligand pathway is a priori unclear. There are also multiple options how transitions between multiple states result in a single on-rate: Either there is a dominant pathway, or multiple pathways with similar rates. In both cases the timescale for the slowest collective process has to be markedly slower than the second slowest process.

What substates are potentially relevant to describe the ligands path into the binding site? The electrostatic and Van-der-Waals interactions, as well as simulations of this system and other protein-ligand-systems (Buch et al., 2011) strongly suggest that ligand binding to the protein surface occurs and plays a role, motivating the postulation of an unknown number of surface states (depicted in black in figure 5.1).

As the ligand somehow has to enter the binding site, the surface states and near-surface states in closest proximity to the binding site have to be visited at some point. Similar to the funnels observed in protein folding, funnels have been postulated and observed (Wang and Verkhivker, 2003; Shan et al., 2011) for ligand binding. Such a funnel corresponds to a set of subsequent, narrowing states, forming a "ladder" into the binding site (depicted in red in figure 5.1). If such a funnel behaves similar to folding funnels, it is accompanied by a free energy barrier that the ligand has to be cross before entering the binding site. Transitions between the different states are governed by according transition probabilities (depicted with black arrows in figure 5.1).

Neither the number nor the shape of the surface states nor the exact transitions are known beforehand. As indicated in section 5.1, there are, however, depending on the weights of the substates and the transition probabilities between the substates, two possible different limiting cases.

In the first limiting case the surface or at least a significant fraction forms a homogeneous state to which the ligand binds from the bulk phase and then enters the binding site via surface rolling. This binding pathway is denoted by the blue and green arrows in figure 5.1. In the second limiting case, the rates from surface states to the binding site are small. Only ligands that attach very closely to the binding site at the surface can actually reach the binding site. This limit then leads to the aforementioned binding funnel.

In the following subsections we will discuss the implications of these binding sites.

5.2.1.1. Three-State Limiting Cases

For the first limiting case, we assume that the binding is determined by a three-state process: A bulk state, a surface state and a binding site state.

In such a model binding occurs via two steps: The ligand first binds via diffusion either to the surface (green in figure 5.1) or a smaller patch of the surface (blue in

figure 5.1) and then rolls on the surface into the binding site:

$$\text{bulk} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{surface} \xrightarrow{k_2} \text{binding site} \tag{5.1}$$

This model describes the binding process accurately either if the rates between the relevant surface states are fast compared to k_1 and k_2 or alternatively if the three rates are similar for all surface states.

In this scenario the rate k_1 for the surface binding depends on the ligand concentration, whereas the surface rolling rate k_2 is concentration independent.

As a result the total effective on-rate behaves differently for varying ligand concentrations: At low concentrations the surface binding is rate limiting and the overall rate appears to depend linearly on the concentration. At high ligand concentrations k_1 is always fast and the concentration independent rates dominate the overall on-rate. As a consequence, depending on the exact rates, the effective on-rate constant becomes concentration dependent.

5.2.1.2. Binding Funnel Case

In the second limiting case the rates from almost all surface states into the binding site are small. Instead, binding occurs almost solely for ligands which bind to states closely to the binding site. In this model the low on-rate is explained by a binding funnel that confines both translational and rotational ligand degrees of freedom.

This confinement is associated with an entropic cost. Since the bound state is known to be stable, there must be an enthalpic attractive part as well that counteracts the entropic confinement. If this enthalpic part is mainly limited to the actual binding site, the combination of enthalpic and entropic parts result in a free energy barrier within the binding funnel for which the uphill part is dominated by the entropic part.

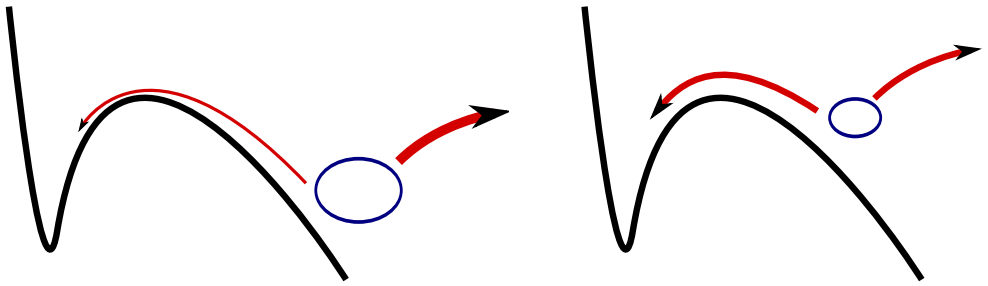
For this model we suggest a technique to derive the resulting on-rate constant from the rotational and translational confinement and the resulting free energy barrier, sketched in figure 5.2a: The on-rate is the product of the rate to reach a translationally and rotationally confined volume (sketched in blue) and the probability to overcome the remaining free energy barrier into the binding site (indicated in the sketch by red arrows).

In this model the on-rate is then quantified by a factor for the translational confinement, the rotational confinement and the barrier crossing probability:

$$k_{\text{on}} = k_{\text{translational confinement}} \cdot p_{\text{rotational confinement}} \cdot p_{\text{barrier crossing}} \tag{5.2}$$

$$= 4\pi N_A D a(x) \cdot \frac{\Omega(x)}{4\pi} \cdot p(\text{bind} \mid x). \tag{5.3}$$

Here we assumed that the translational confinement within the binding funnel is described by a Smoluchowski term $4\pi N_A D a(x)$, where D is the diffusion coefficient



(a) (b)

Figure 5.2.: In the funnel model, the on-rate is given by diffusion towards are translationally and rotationally confined volume (blue) near the binding site, combined with the probability to cross the remaining free energy barrier. If the barrier is mainly determined by the confinement of the configurational space, i.e. mainly entropic, the on-rate estimate does not depend on the choice of the position of the confined volume: The probability to reach the target volume in (a) is larger than in (b), but the probability to overcome the remaining barrier is smaller in (a) than in (b).

of the ligand and $a(x)$ describes the effective radius of the target volume at a given distance x from the binding site along a suitable reaction coordinate (i.e. the radius of the blue circle in figure 5.2a).

The effect of rotational confinement depends on the rotational diffusion: Analytical calculations for reaction kinetic between spherical and axially symmetric but chemically non-uniform molecules (Šolc and Stockmayer, 1971) show that surface heterogeneity leads to an effective rate constant which differs from Smoluchowski's diffusion controlled rate constant by factors smaller than unity. Their quasi-chemical approximation (Šolc and Stockmayer, 1973) indicates that the rotational confinement is only relevant for slow rotational diffusion. Since the smallest moment of inertia of cAMP is 6-7 times smaller than the larger two, we assume that rotation around the main ligand axis is comparably fast and can thus according to Šolc and Stockmayer (1973) be neglected, whereas confinement of the main axis orientation is relevant.

As a consequence, in our model the rotational confinement is given by the solid angle $\Omega(x)$ available to the ligand at a given point in the binding funnel

If the enthalpic contribution to the uphill part of the barrier is small, the on-rate estimate does not depend on the choice of the position of the confined volume, i.e., on the point x in equation 5.3 (see figures 5.2a and 5.2b): The slower rate to reach a more confined volume is exactly compensated by the increased probability to cross the barrier.

5.2.2. Markov Model of Ligand Binding

A priori neither the number nor the shape of the surface states sketched in figure 5.1 are known. The identification of binding and surface substates and the determination of transition times between them is however a nontrivial task. Markov models (Prinz et al., 2011; Weber, 2012; Buch et al., 2011) are an effective tool to analyse MD trajectories to determine both the equilibrium distribution as well as the slowest transitions and the corresponding timescales. In principle the states in figure 5.1 already constitute a discretisation for a Markov model. Their exact shape, however, are unknown. We therefore constructed a Markov model of the ligand dynamics with a finer discretisation. The larger substates sketched in figure 5.1 then simply consists out of multiple substates from the detailed model.

In a Markov model, the configurational space is discretised into states and the dynamics is given by transitions between the states. A Markov model therefore consists of the discretisation and the set of transition probabilities between these states.

For the construction of Markov models the configurational space is discretised into small bins. Subsequently transitions occurring in a lag time τ between these bins are counted from the discretised trajectories. We denote the number of counted transitions between states i and j as c_{ij}. The whole set of transition counts thus forms a matrix $\mathbf{C}$. The pure count matrix is then transformed into a transition matrix $\mathbf{T}$, where the elements t_{ij} denote the probability of a transition from i to j within the lag time τ under the condition that the system is already in state i. The transition probabilities t_{ij} of the transition matrix are estimated via

$$t_{ij} = \frac{c_{ij}}{\sum_j c_{ij}}. \tag{5.4}$$

The eigenvectors of the transition matrix correspond to a probability flow from a subset of discretisation cells into other cells and vice versa. The corresponding eigenvalues directly translate into the timescales of the associated probability flow. If the underlying dynamics is ergodic, the first and only the first eigenvalue is 1, which corresponds to an infinite timescale. Consequently, the first eigenvector corresponds to the equilibrium probability distribution and therefore plays a distinguished role.

With the determination of the first eigenvector, a Markov model offers a way to determine the equilibrium distribution from a set of non-equilibrated simulation trajectories. We used this feature to determine which states in figure 5.1 are populated in equilibrium.

All other eigenvalues are smaller than 1 and thus translate into finite timescales, in particular the second eigenvector thus corresponds to the slowest global process in the model and the higher eigenvectors correspond to faster processes.

The MD simulations include all degrees of freedom of ligand, protein and solvent. For the identification of ligand binding substates however, we are mainly interested in the ligand position and conformation. Therefore we only used the ligand's COM

and orientation for the construction of the Markov model. This reduction of the conformational space impairs the Markovianity of the resulting model: Iterated n-times multiplication of the transition matrix $\mathbf{T}(\tau)$ for a lag time τ with a probability vector $\mathbf{p}_0$ does not yield the same result as a transition matrix $\mathbf{T}(n\tau)$ built for the lag time $n\tau$:

$$\mathbf{p}_0^T \mathbf{T}(\tau)^n \neq \mathbf{p}_0^T \mathbf{T}(n\tau). \tag{5.5}$$

Compared to an exact model, timescales are underestimated, including the timescales of the dynamics described by the first eigenvectors. The error cannot be estimated beforehand, however, it has been shown that this error in the Markov property decreases if the Markov model is built for larger lag times τ (Prinz et al., 2011), allowing for the recovery of slow timescale transitions and especially the stationary solution.

Apart from considering only the ligands degrees of freedom, the discretisation of these ligand degrees of freedom, in our case the ligand COM position and its orientation, introduces an additional discretisation error (Prinz et al., 2011) which additionally impairs the Markovianity of the model.

To handle these issues and to obtain an estimate for the slowest timescales, different discretisations were performed. As the error decreases with finer spatial discretisation and coarser temporal discretisation (i.e., a larger lag time) we are able to extrapolate to the actual timescales of the ligand dynamics in the MD simulations.

5.3. Methods

5.3.1. Molecular Dynamics Simulations

5.3.1.1. Simulation Parameters

The employed simulation protocol, including the parametrisation of the ligand, is identical to the one for the unbiased simulations in the previous chapter and therefore already described in section 4.3.1, except that cubic simulations boxes were used for the simulations carried out in this chapter.

5.3.1.2. Modelling of Protein Structure

All ligand binding simulations at the CNBD were based on the R384A mutant (PDB ID code 1U12) (Clayton et al., 2004) of the CNBD. The mutation was reverted by replacing residue Ala384 with Arg384 and missing atoms were added using the Modeller 9v10 software (Šali and Blundell, 1993).

5.3.1.3. High-Concentration Simulations

900 200 ns simulations of dissolved cAMP and CNBD were carried out. For the starting configuration for every simulation setup a single cAMP molecule was randomly rotated and placed in a cubic box with a box length of 7 nm together with the open conformation CNBD. The box size corresponds to a concentration of 4.84 mmol/l. Sodium chloride ions were added at a concentration of 113 mmol/l.

5.3.1.4. Low-Concentration Simulations

For cross-validation simulations, 100 300 ns simulations and 100 200 ns simulations were carried out in a cubic box with box length of 14 nm, corresponding to a concentration of 0.61 mmol/l. As a starting configuration, ligand molecules were placed in random orientation spherically in a distance of 6.5 nm around the centre of protein. Sodium chloride ions were added at a concentration of 113 mmol/l.

5.3.2. Data Analysis

5.3.2.1. Markov State Model

The set of high concentration simulations was used to construct multiple Markov models of the ligand motion.

Discretisation All ligand trajectories were fitted with reference to the CNBD orientation. Due to the cubic box shape and the CNBD rotation during the simulations, the space of ligand positions with respect to the protein is larger than the simulations box, its volume is $V = (\sqrt{3}a)^3 = 3V_0$. For different Markov models, this volume was discretised into 10^3, 20^3 or 30^3 cubic bins. For different Markov models, the rotational degrees of freedom was discretised into 1, 6 or 36 bins. To do that only the first or the first and second main axis were projected onto cube, yielding 6 different bins per rotational discretisation. The resulting discretisations thus consisted of 1000 (only spatial) to 972000 (27000 spatial times 36 rotational) bins.

Count Matrix Calculation Count and transition matrices were calculated for all discretisations and various lag times τ ranging from 1 ns to 80 ns. Transition matrices were calculated directly from the count matrices via $T_{ij} = \frac{C_{ij}}{\sum_j C_{ij}}$ and using an algorithm by (Prinz et al., 2011) based on (Bowman et al., 2009) that yields a maximum probability estimator for a reversible transition matrix.

Eigenvalues λ_i and eigenvectors v_i were calculated for all transition matrices, as well as implied time scales via $t_i = \frac{\tau}{\ln \lambda_i}$.

5.3.2.2. Comparing Surface Binding with Diffusion

To test whether the surface binding from bulk phase is controlled by diffusion, we determined for the low concentration MD simulations the number of trajectories where no direct ligand protein surface contact occurred until certain fixed times. Subsequently the theoretically expected numbers were determined. To that end the probability that no surface binding occurs was calculated using equation 2.74:

$$p(t) = \exp\left[-\frac{4\pi DR}{V} \cdot t \left(1 + \frac{2R}{\sqrt{\pi Dt}}\right)\right]. \tag{5.6}$$

Here $D = 1.42 \cdot 10^{-9}\,\mathrm{m^2/s}$ is the diffusion coefficient of cAMP in TIP3P water during the MD simulations, $R = 2.2\,\mathrm{nm}$ the approximate radius of the protein and $V = 2.744 \cdot 10^{-24}\,\mathrm{nm^3}$ the simulation box volume. The expected number of non-surface-binding trajectories out of a total of $N = 198$ simulations was then

$$\langle n \rangle(t) = N \cdot p(t) \tag{5.7}$$

$$\sigma_n(t) = \sqrt{N \cdot p(t) \cdot (1 - p(t))}. \tag{5.8}$$

5.3.2.3. Rate Extrapolation in the Three-state Model

Rates in the three-state model were calculated by setting up a 3-state transition matrices representing transitions between bulk, surface and binding site states for different ligand concentrations. The mean first passage time was calculated using the group general inverse (Meyer, 1975). The effective rate was calculated as the inverse of the mean first passage time.

Surface Target Area Calculation To calculate the area of the protein surface on which the ligand attach before entering the binding site, the following technique was employed: For all surface attaching structures the coordinates of the atoms were determined using the COM of the protein as the origin. All atom coordinate vectors were then normalised to a length of 2.2 nm, corresponding to the approximate protein surface sphere radius. Subsequently 10000 uniformly distributed random points on the surface were drawn. For each point it was checked if it was within a distance of less than 0.12 nm from an atom point. The fraction of random points within the 0.12 nm radius of the atom points was interpreted as the surface area fraction to which the binding molecules attached.

5.3.2.4. Binding Funnel Parameters

To describe the progress of the ligand to the binding site we choose the ligand root mean square deviation (RMSD) to the bound state configuration as a reaction coordinate. The bound state configuration was extracted from the x-ray structure of the cAMP bound CNBD (PDB code 1VP6) (Clayton et al., 2004).

Diffusion Term To estimate the radius of the effective target sphere, we identified conformations that were almost visited by every other trajectory were successful ligand binding occurred.

To that end for every ligand conformation in the ligand binding trajectories the RMSD and the COM were calculated as well as for every conformation from every ligand binding trajectory the conformation that had the closest RMSD to the given structure. From this ensemble of ligand conformations the maximal COM distance was calculated, which was used as the estimate for the diameter of that particular "neighbourhood cloud".

In a second step the minimal neighbourhood cloud diameter for every RMSD value was calculated, yielding the functional dependence of the target sphere radius a on the ligand RMSD to the binding site x, $a(x)$.

Solid Angle Term The rotational confinement was estimated from the scattering of the major orientation axis of the ligand. We calculated for every RMSD to the ligand binding site the ensemble of structures with that RMSD from the ligand binding trajectories and estimated the solid angle opened by these conformations.

To that end for every value along the reaction coordinate all structures from the ligand binding trajectories with the corresponding reaction coordinate value were extracted. The vector defined by the phosphor atom and a nitrogen atom at the opposite side of the molecule was taken to represent the major axis of the molecule. Of the resulting ensemble of vectors we calculated the central vector and the angle between all vectors and the most central vector, yielding a distribution of angles. The half apex angle θ of the cone was the, estimated as the mean of this angle distribution plus one standard deviation of the angle distribution. The solid angle or more precisely the fraction of a full sphere is then

$$\Omega(x) = \sin^2\left(\frac{\theta}{4}\right) \tag{5.9}$$

$$= \sin^2\left(\frac{\mathrm{E}(\mathrm{angledist}(r)) + \sqrt{\mathrm{Var}(\mathrm{angledist}(x))}}{2}\right). \tag{5.10}$$

The same procedure was applied to the whole set of trajectories (i.e. without filtering for trajectories where binding actually occurs).

Rotational Confinement of the Protein The target sphere is located on the protein surface. Therefore it is only accessible from one side. This confinement, which corresponds to a rotational confinement of the protein, is taken into account by introducing a factor of 0.5 into the rotational confinement.

Barrier Crossing Term The probability of binding (before unbinding occurs) for a given point along the ligand binding reaction coordinate was calculated straightforward from the binding trajectories. Unbinding was defined by crossing an upper

cutoff of 1.0 nm. Binding was defined by reaching a lower cutoff. Multiple lower cutoff values between 0.1 nm and 0.6 nm were used.

For every value along the reaction coordinate between lower and upper cutoff and for all trajectories it was determined if and when this reaction coordinate values were reached. Subsequently it was observed whether the ligands RMSD became smaller then the lower cutoff, larger than the upper cutoff or did not escape this interval. All events were counted, yielding a probability for binding (i.e reaching the lower cutoff) before unbinding under the condition of having already reached a certain point x along the reaction coordinate: $p(\text{bind} \mid \text{RMSD} = x)$.

5.3.2.5. Brownian Dynamics in a One-Dimensional Model Landscape

Comparison of commitor functions from the MD simulation data with a one-dimensional barrier were performed via Brownian dynamics simulations in a simple toy model.

To that end Brownian dynamics trajectories in an inverse harmonic potentials of the shape $U = -\frac{\alpha}{2}(x - x_r)^2$ were carried out. The overdamped stochastic dynamics equation of motion were integrated using the Euler-Maruyama method.

Trajectories were stopped whenever $x(t)$ became larger than 1 nm, which is the upper cutoff used for the determination of the commitor functions, or smaller than a lower cutoff. Starting positions x_0 were varied between the lower and upper cutoff and 5000 simulations were carried out per starting position. This procedure was repeated for 11 different values for the lower cutoff between 0.1 nm and 0.6 nm.

Subsequently for every lower cutoff and starting position, the ratio of the number of times the lower cutoff was reached and the number of times the upper cutoff was reached was calculated. To score how similar these values were to the ratio of the commitor functions from the MD simulations, the following measure was employed:

$$\varepsilon(\alpha, x_r) = \sum_{\text{lower cutoff}} \sum_{\text{start point}} \left[\log\left(\frac{n_{\text{up}}}{n_{\text{low}}}\right) - \log\left(\frac{p_{\text{unbind}}}{p_{\text{bind}}}\right) \right]. \tag{5.11}$$

$\varepsilon(\alpha, x_r)$ was calculated for different force constants α between $5\,{}^{\text{kJ}}/_{\text{mol·nm}^2}$ and $30\,{}^{\text{kJ}}/_{\text{mol·nm}^2}$ and different maxima x_r between 0.2 nm and 0.8 nm. Finally the pair α, x_r yielding the smallest ε was determined.

5.4. Results and Discussion

5.4.1. Markov Model

First we want to address the question what states exist, where they are located and what transitions between these states occur. For that reason we discretised the set of trajectories where unbound cAMP is simulated in solution with ligand-free

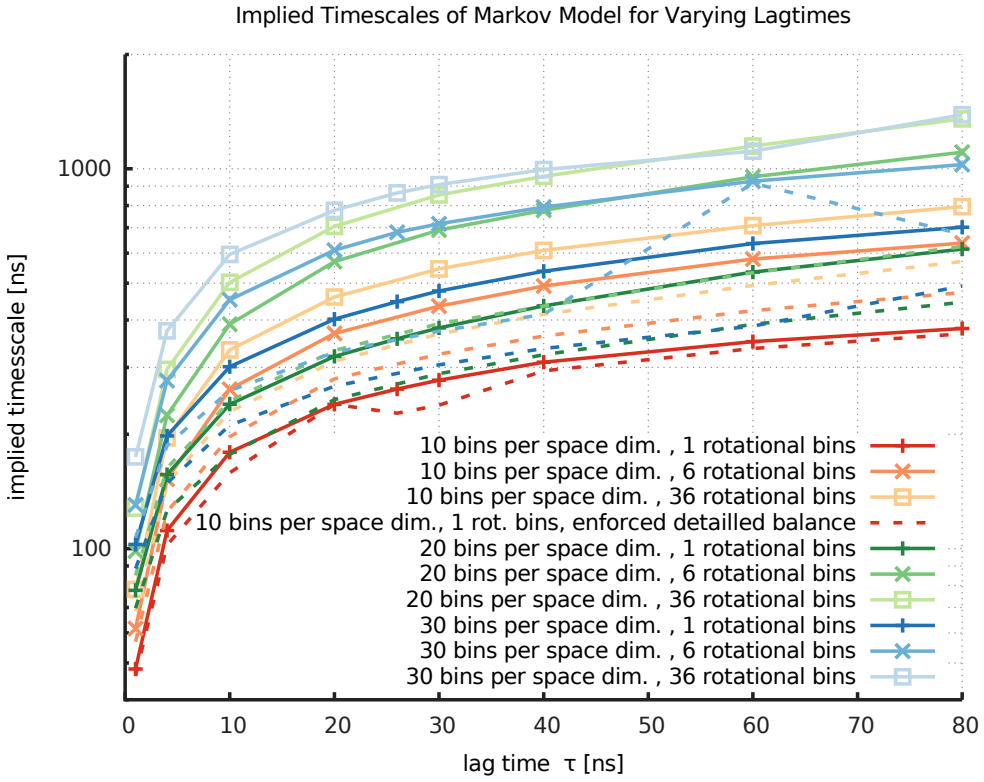


Figure 5.3.: Implied timescales of the slowest process for the ligand binding Markov model in dependence of the lag time τ used to build the Markov model. Different colours show different spatial and rotational discretisations of the ligand dynamics. Dotted lines are implied timescales for Markov models were detailed balance was enforced.
The implied timescale converge only slowly and depend on the exact discretisation.

CNBD and built a Markov model (Prinz et al., 2011; Weber, 2012; Buch et al., 2011) from the discretised trajectories.

The discretisation was done using the ligand COM position and orientation. To assess the quality of the Markov model, multiple voxel size for the spatial bin sizes were tested, namely 1.2 nm, 0.6 nm, 0.4 nm and 0.3 nm and different number of rotational bins, namely 0 (no rotational discretisation), 6 and 36 rotational bins (see sections 5.2.2, 5.3.1.3 and 5.3.2.1).

5.4.1.1. Implied Timescales of Slowest Processes

The first task is to assess the quality of the obtained Markov Model i.e. the effects of the error imposed by the finite discretisation upon the Markov property of the model. Since the discretisation error decreases with increasing lag time and increasing spatial resolution, the implied timescale of the slowest process for

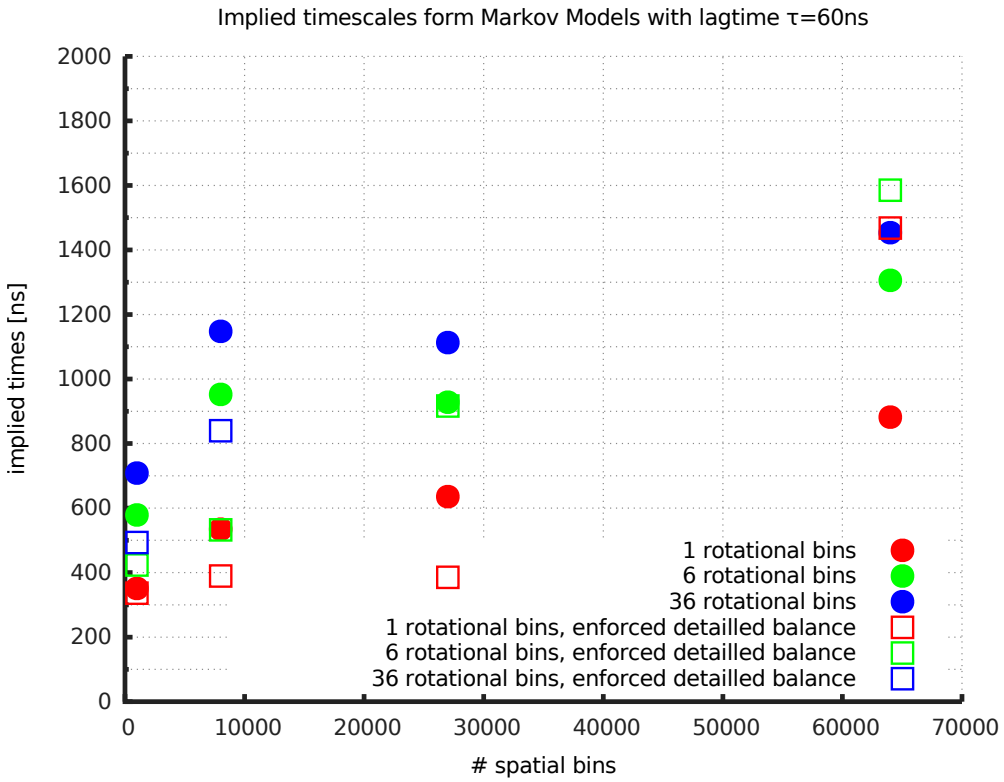


Figure 5.4.: Implied timescales of the slowest process for the ligand binding Markov model in dependence of the spatial and rotational discretisation for a fixed lag time of $\tau = 60\,\mathrm{ns}$. The implied timescales seem to converge with finer discretisations below a value of $2\,\mu\mathrm{s}$.

different temporal and spatial discretisations is calculated. Figure 5.3 shows that the prediction for the intrinsic time scale increases with increasing lag time and converges only slowly. It furthermore depends strongly on the chosen discretisation: As expected, a finer spatial and rotational discretisation leads to more accurate results. Figure 5.4 furthermore shows that the predicted intrinsic time scale for a fixed lag time also converges with increasing number of discretisation bins.

Although the Markov model suffers from discretisation errors, we are able to roughly extrapolate at which value the implied timescale of the slowest process would converge. We therefore estimate the time scale of the slowest process to be between 1 and 2 µs. The time scales for the third and fourth eigenvector are also roughly 1 µs, and the one of the forth eigenvector are about 500 ns and less.

5.4.1.2. Eigenvectors

The first five eigenvectors are shown in figure 5.5. As expected, the first eigenvector, the stationary probability distribution (figure 5.5a), shows a peak inside the binding pocket. However, it also reveals several additional binding spots on the surface of the protein.

This means that the postulated surface states do exist and are populated.

The higher eigenvectors (figure 5.5b-5.5e) describe the probability flow of the processes with the slowest time scales in descending order. The slowest process, shown by the second eigenvector, depicts the probability flow between the binding site and the rest of the system.

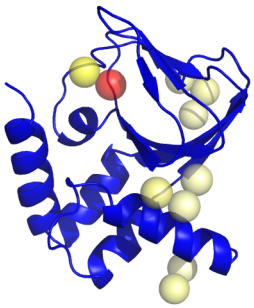
The second, third and fourth eigenvector correspond to transitions solely between different regions on the surface of the protein. Remarkably, these vectors do not have a significant amplitude in the region of the binding site. Instead the describe transitions solely between different surface states. Their corresponding timescales are between 1 µs and 500 ns, i.e. in a similar regime as the binding to the binding site.

These vectors clearly show that for the given concentration the ligand binding process cannot simply be described by an unbound solution state and a bound state. In fact several surface-bound states influence the dynamics.

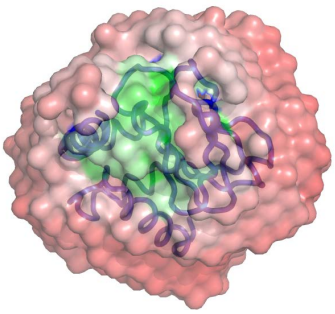
A Markov model that correctly predicts the slowest time scales of the ligand dynamics can only be constructed for lag times of 80 ns and longer. However, the average time until a ligand in bulk phase hits the protein surface for the given simulation box size is under 10 ns. As a consequence, it is from the Markov model unclear whether the transitions between the shown sub states are due to surface rolling or surface detachment followed by attachment at another surface spot.

5.4.2. Protein Surface Attachment

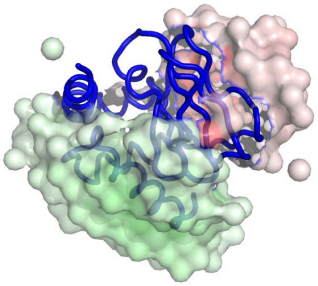
To resolve this issue, we determined the points from where the ligand can roll into the binding site along the surface without further detachment. These points are defined as the points where the ligand, in trajectories in which successful binding



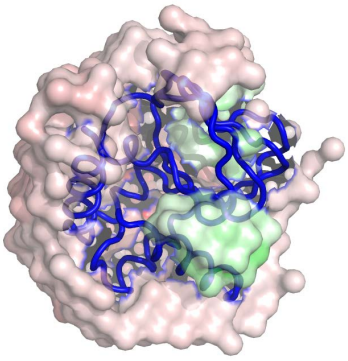
(a) 1st Eigenvector: stationary distribution



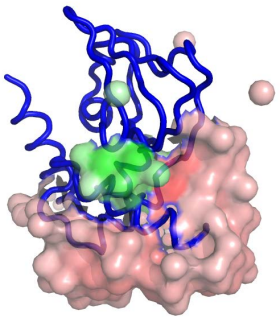
(b) 2nd Eigenvector



(c) 3rd Eigenvector

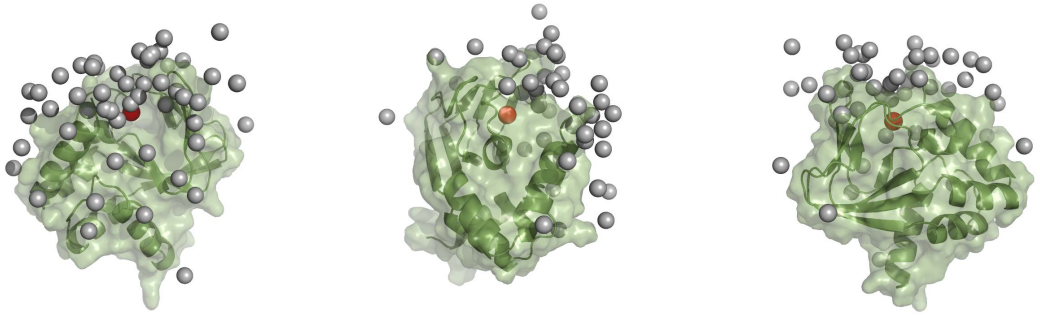


(d) 4th Eigenvector



(e) 5th Eigenvector

Figure 5.5.: Eigenvectors of the Markov model of the ligand binding process. The 1st left eigenvector of the transition matrix constitutes the stationary distribution. All higher eigenvectors reveal the probability flow of the slowest transitions, with the 2nd eigenvector being the slowest one. Here the right eigenvectors of the transition matrix are depicted, that show the probability flow unweighed by the stationary probability distribution.



(a) front view (b) side view (c) back view

Figure 5.6.: From where can ligands roll into the binding site? Grey spheres show where ligand molecules attach to the surface in ligand binding trajectories before rolling on the surface into the binding site. The binding site is shown as a red sphere. Only ligands that attach to the surface sufficiently close to the biding site can actually bind before detaching again.

occurred, first attaches to the surface and does not detach again before entering the binding site.

If these points are distributed over the entire protein surface, then transitions into the binding site are possible from all surface states shown before. Furthermore transitions between the surface states should also be possible via surface rolling. That means the ligand's path into the binding site would be described by the green pathway in figure 5.1. If the surface attachment points only occupy a part of the protein surface the binding path is better sketched by the blue pathways in figure 5.1. If the attaching points are mainly located around the binding site itself the funnel model is most likely.

The surface attaching conformations, shown in figure 5.6 reveal that only ligands binding in an area around the binding site can actually roll into the binding site. This part constitutes 0.3 ± 0.05 of the total protein surface (see method section 5.3.2.3). In most trajectories, however, the final surface attaching occurs very close to the binding site though.

Vice versa this result directly excludes the possibility that the ligand only has to bind to the surface and then rolls into the binding site (sketched in green in fig. 5.1). Since the ligand apparently does not roll over the entire protein surface, the slowest transitions seen in the eigenvectors of the Markov model (figure 5.5) probably also consist of detachments into the bulk phase and reattachments to other surface states. Since the diffusion within the bulk phase is fast for the used simulation box size, the microsecond transition times between the surface binding spots are most likely dominated by large off-rates from the surface binding states.

If now the transitions shown by the eigenvectors of the Markov model consist of detachments, bulk diffusion and reattachment, then it is to be expected that

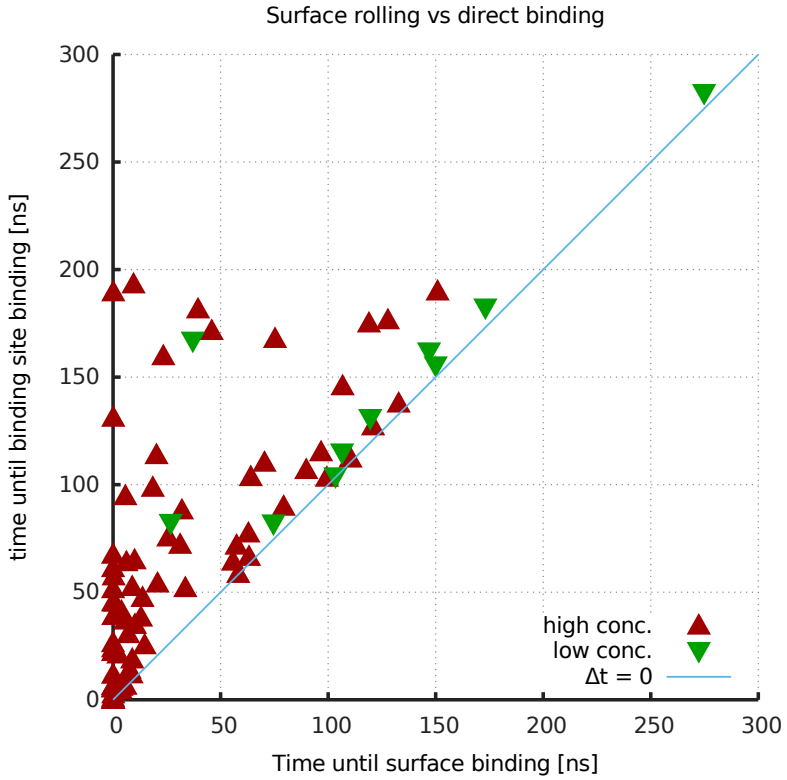


Figure 5.7.: Time until binding to the binding site occurs (y-axis) vs. time until final surface attachment occurs (x-axis) in trajectories where ligand binding actually occurs. In many trajectories binding to the binding site occurs shortly after the final surface attachment, showing that surface attachment occurs already close to the binding site.

in larger simulation boxes (i.e. lower concentrations) these transitions would take longer. The timescales from the Markov model as such is only valid for the fixed ligand concentration.

Besides the spatial informations, i.e., *where* ligands that bind attach to the protein surface, we also investigated the timescales of the transitions. To that end, we measured how long the ligands needed to reach the binding site from the surface attachment points shown in figure 5.6 and compared it to the times needed until this final surface attachment occurred (figure 5.7). Apart from simulations where the ligand is already bound at the beginning of the simulation, in most cases binding to the actual binding site and surface attaching happened within a short time frame. Together with the observation from the Markov model, that transitions between distant surface binding patches are comparably slow, this finding underlines that the area from where ligands effectively roll into the binding site is limited in size around the actual binding site.

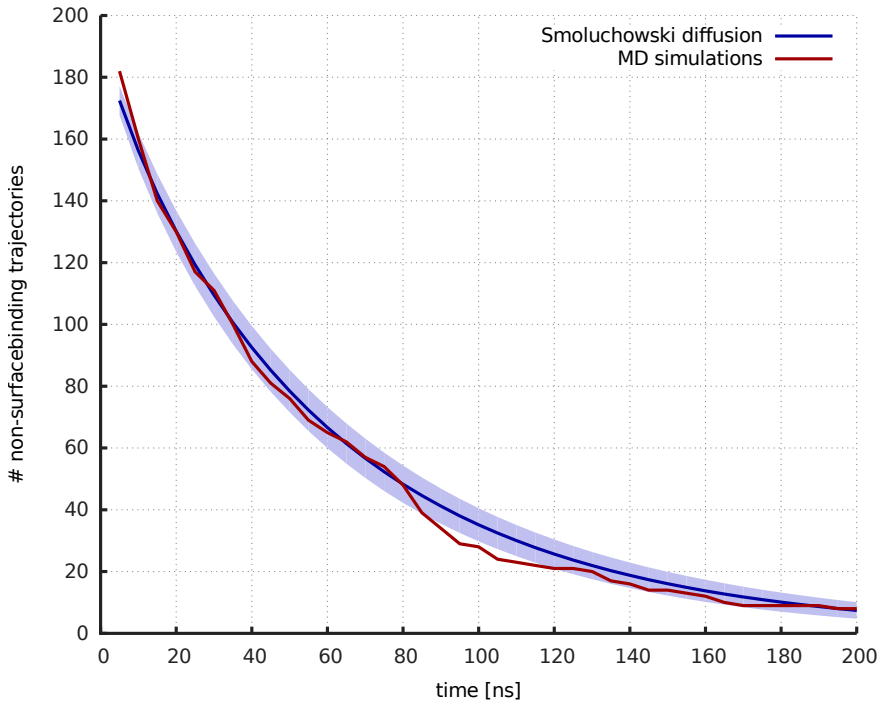


Figure 5.8.: Surface binding is limited by diffusion: Number of trajectories in low concentration simulations where, until a given time, no surface-contact occurred. The red curve shows data from actual MD simulations. The blue curve shows the number expected from pure Smoluchowski diffusion. A total of 200 simulations were carried out.

5.4.3. Three-State Model

In the limiting case where the full model is reduced to a three-state model the ligand first binds to the protein surface or a part of the surface and subsequently rolls into the binding site or detaches again.

First we investigated if the rate of surface attaching is actually fully determined by diffusion. To that end we determined from the low concentration simulations (see section 5.3.1.4) the number of trajectories where no surface binding occurs and compared this with the expected number from equation 2.74. The result is shown is figure 5.8. The curves agree well within the statistical errors of the theoretical curve. It should be noted that this agreement only holds for the exact solution (equation 2.74), whereas the approximate solution for ong timescales (equation 2.75) significantly deviates from the observed curve.

This nice agreement shows that the binding from the bulk phase to the protein is completely controlled by diffusion.

The diffusion coefficient however, calculated from the mean displacement in solution with time, differs by a factor of approximately 3 from the experimental value of $D = 0.44\,\mathrm{m^2/s}$ (Dworkin and Keller, 1977). This discrepancy can be attributed

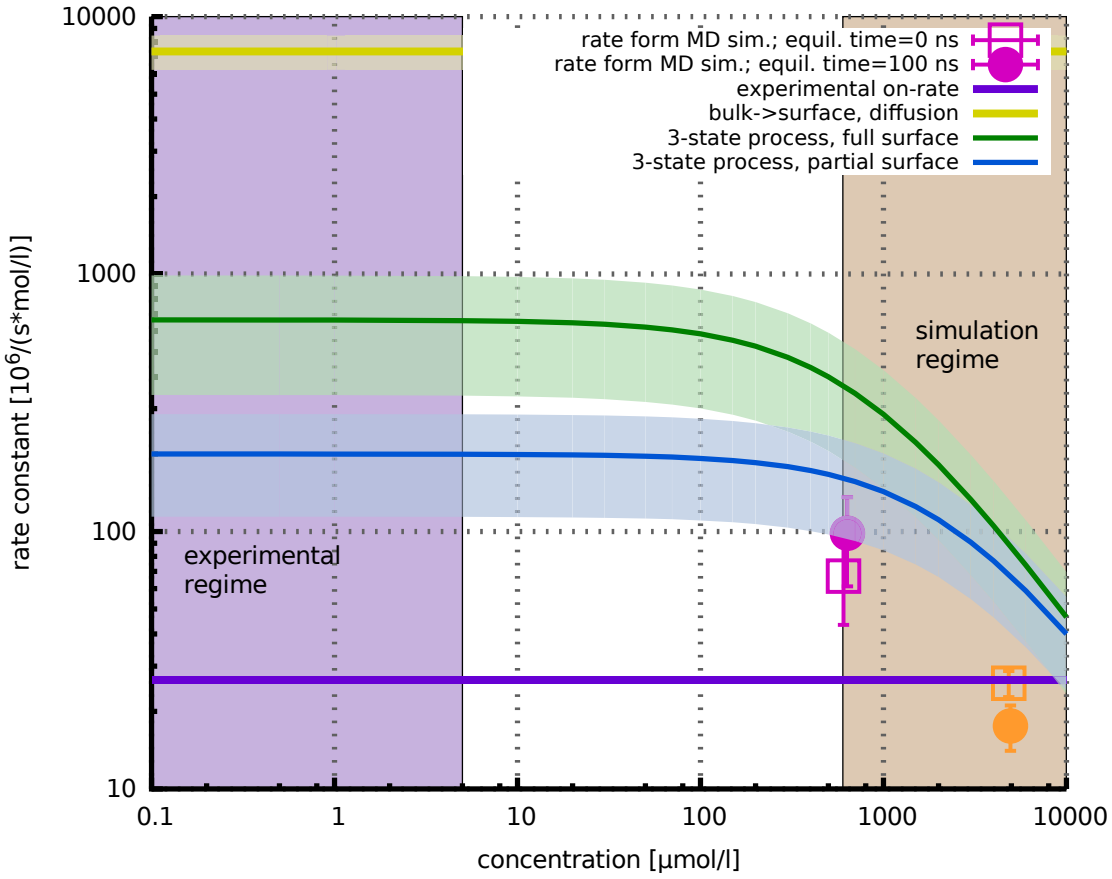


Figure 5.9.: Hypothetical on-rate constants in various models in dependence of the ligand concentration. Yellow: Pure diffusion to the protein surface. Green: on-rate if the binding were given by a three-state process: bulk↔surface↔binding site. Blue: Same as green, but considering only part of the surface. Violet: experimentally determined on-rate constant. The rate for surface binding from the bulk phase is always assumed to be diffusion limited, the hypothetical surface rolling rate and unbinding rate are taken from the Markov model.
The rate for surface rolling does not depend on the ligand concentration, the diffusive surface binding however does. Therefore the rate constant is independent of the concentration for low ligand concentrations and ligand concentration dependent for high ligand concentrations.

to the TIP3P water model, which is known to yield too large diffusion coefficients (Mark and Nilsson, 2001; Mahoney and Jorgensen, 2001; Patra and Karttunen, 2004). The discrepancy is corrected for in all following considerations.

To test if a three-state model can explain the low measured on-rate, we determined the effective on-rate constant for different ligand concentrations, ranging from the low concentrations corresponding to the experimental conditions until the high ligand concentrations corresponding to the conditions in the MD simulations (fig. 5.9). We used Smoluchowski's on-rate constant with the experimental diffusion coefficient for the transition from bulk to surface. The rates for the surface off-rate and the rate for transitions from the surface to the binding site were extracted from the Markov model.

A naive estimate for the on-rate directly from the MD simulations was obtained via $k = n/t$, where n is the number of binding events and t the accumulated times until binding occurs or the trajectory ends. While these rate estimates fit roughly both to the experimental value and the rate constant predicted by the three-state model, it is apparent that for low concentrations the three-state model predicts an on-rate that is significantly larger than the measured experimental value.

As a result, we can exclude the three-state model, i.e. the green and blue pathways from figure 5.1. The most likely reason is that the assumption of a homogeneous surface state is flawed.

5.4.4. Binding Funnel

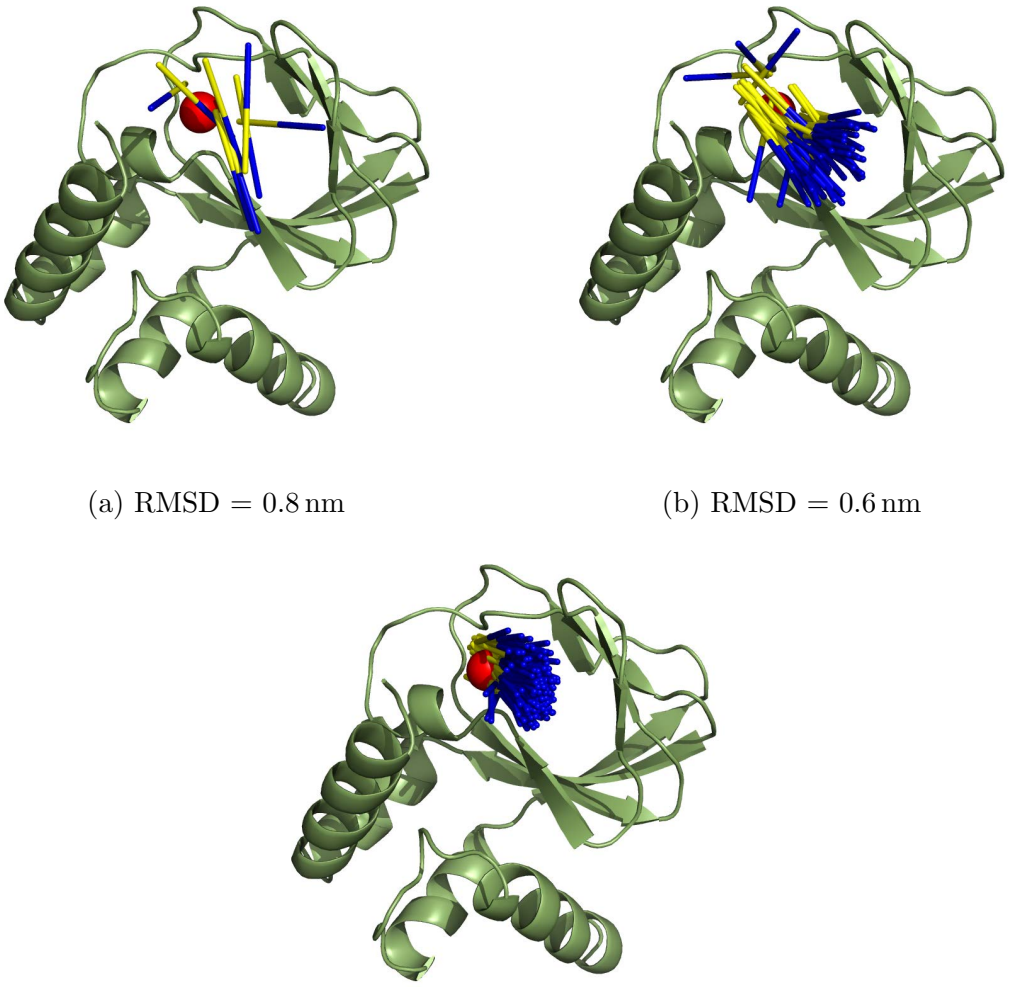
The alternative limiting case is that there is no clear intermediate binding step, but that the ligand has to diffuse into a binding funnel. A natural measure of the ligand's progress towards the binding site is the RMSD to the bound conformation, which shall in the following serve as a reaction coordinate.

A visual inspection of ligand ensembles with identical RMSD to the bound configuration (figure 5.10) already shows that the conformations in the ensembles become more and more similar with decreasing RMSD. Vice versa this observation implies that binding requires translational and rotational confinement, motivating the funnel model.

In the following this confinement is quantified, thus enabling us to estimate the on-rate using equation 5.3.

5.4.4.1. Translational Confinement

To calculate the translational confinement, for every ligand conformation the most similar conformation of every other binding trajectory is determined. The radius of the ensemble of COMs of these ligand configurations, which we refer to as the *neighbourhood cloud radius*, then determines how close a binding trajectory has get to the given conformation (see section 5.3.2.4). The minimal neighbourhood cloud radius in dependence of the reaction coordinate is shown in figure 5.11. The radius of the neighbourhood cloud monotonously decreases with decreasing RMSD



(a) RMSD = 0.8 nm

(b) RMSD = 0.6 nm

(c) RMSD = 0.3 nm

Figure 5.10.: Ligand ensembles from different binding trajectories for different ligand RMSDs to the binding site. The ligand conformations are drawn as sticks, showing their main orientation. The COM of the bound ligand conformation is shown in red. With decreasing RMSD, the ligands conformational flexibility is increasingly confined, leading to a more homogeneous ensemble.

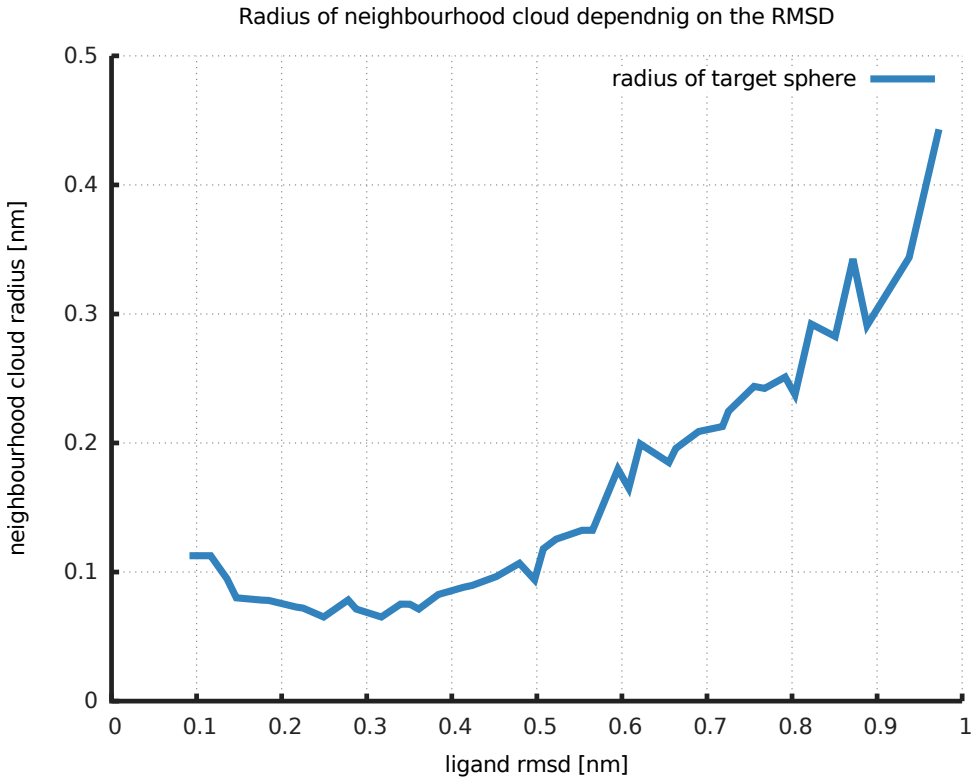


Figure 5.11.: Radius of the "neighbourhood cloud" in dependence of the ligand RMSD to the binding site. The radius of the neighbourhood cloud is a measure of the volume to which the COMs of the ligand molecules are confined in ligand binding trajectories. The accessible volume decreases the closer the ligand gets to the binding site, showing and quantifying the increasing translational confinement.

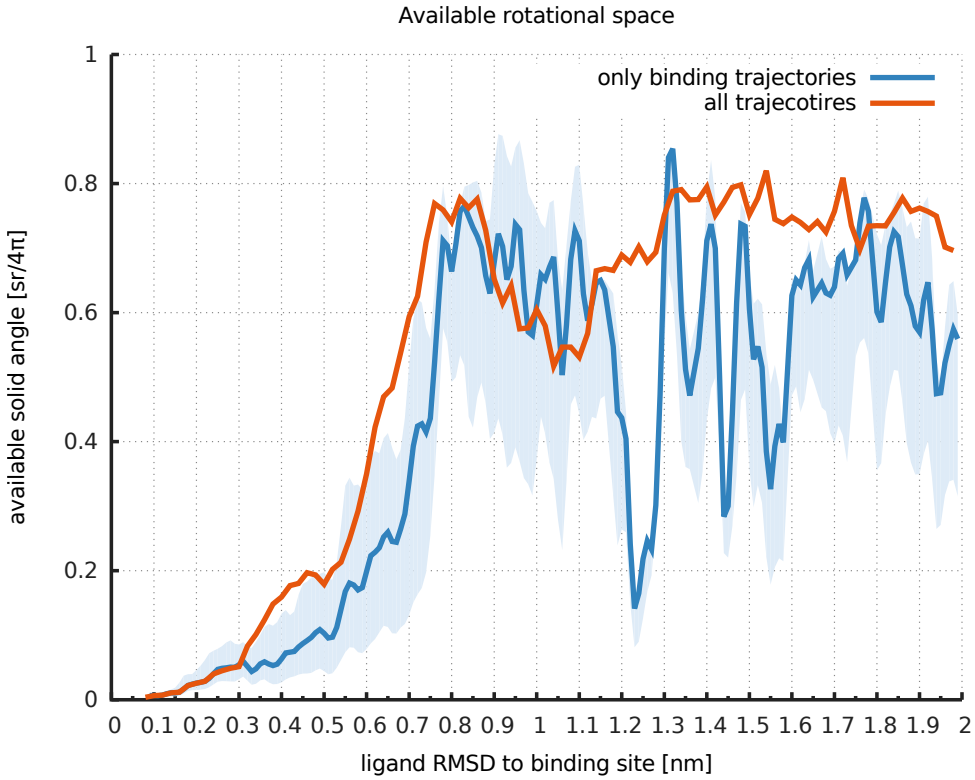


Figure 5.12.: Available rotational space, measured by the estimated solid angle formed by the major axes of an ensemble of ligand conformations. A value of 1 means no rotational restriction. As soon as the ligand enters the binding funnel (at approximately 0.8 nm), the rotational degrees of freedom are drastically decreased.

to the binding site, which is in excellent agreement with the picture of a narrowing binding funnel.

The ligand is already well bound for an RMSD of 0.3 nm, thus the small rise in the neighbourhood cloud radius for very small RMSDs can be attributed to the limited sampling in this regime.

The neighbourhood cloud radius is then identified as the target radius $a(x)$ from equation 5.3.

5.4.4.2. Rotational Confinement

Besides the COM, the orientational degrees of freedom of the ligand are also reduced as the ligand enters the binding funnel. To quantify this, we grouped the structures from trajectories where binding occurs according to their RMSD. For every ensemble of conformations, the available solid angle spanned by the vectors defined by their major axis was calculated (see method section 5.3.2.4 for details). We thereby obtained a measure for the rotational restriction (figure 5.12). For an

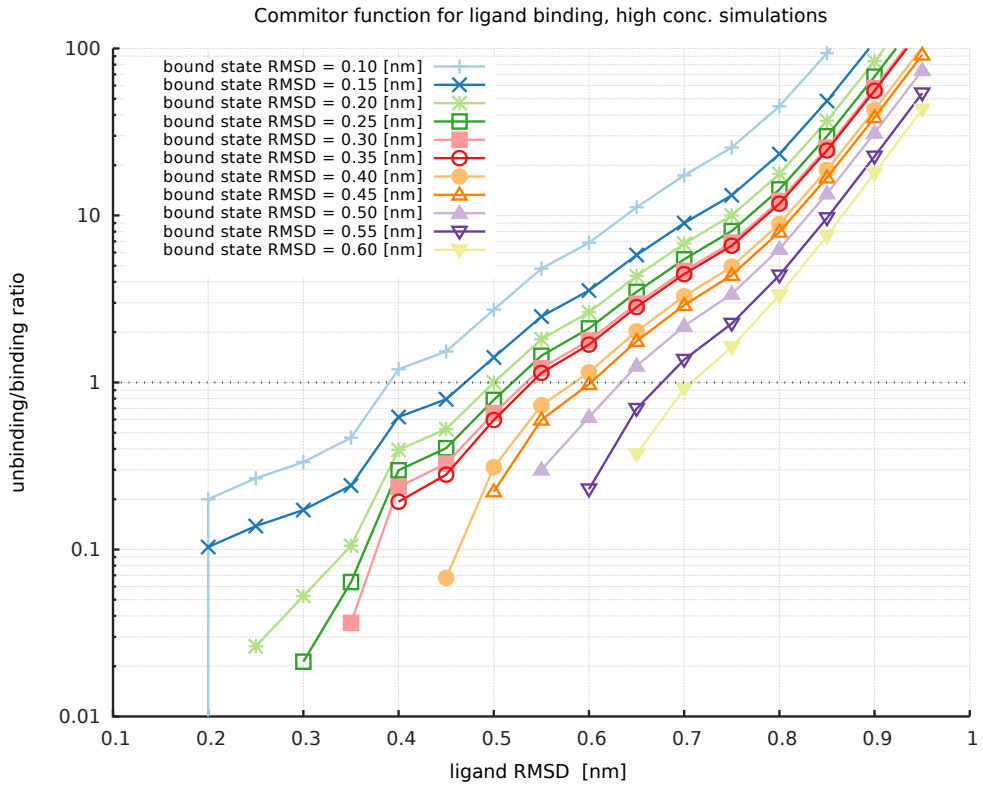


Figure 5.13.: Ratio of committor functions for ligand unbinding and ligand binding. The commitor functions are the probability that the ligand reaches a certain lower (or upper) cutoff given that is already has a certain RMSD to the binding site. A committor ratio of 1 corresponds to the transition state where unbinding and binding are equally likely. Different colours represent different values for the lower cutoff, i.e. different possible points where in a experimental setup fluorescence of a labelled ligand could start.

RMSD of larger than 0.8 nm the ligand is rather free to assume arbitrary rotations, for smaller RMSD values however, the available rotational space is increasingly confined.

It is furthermore noteworthy that the available solid angle, calculated for all trajectories is especially at the funnel entrance, i.e. at a ligand binding site RMSD of around 0.7 nm, much larger than in trajectories where binding occurs. In this light it is plausible that successful binding is dependent on the ligand orientation at the funnel entrance.

5.4.4.3. Barrier Crossing Probability

The ligand can dissociate from the binding funnel at every point along the reaction coordinate. We therefore calculated the conditional probability that the ligand

actually reaches the binding site given a certain point along the reaction coordinate. We then combined this binding probability with the calculated translational and rotational confinement to obtain an on-rate estimate according to equation 5.3.

To that end we calculate for all trajectories that assume values along the reaction coordinate the commitor function, i.e. the probability to reach a lower cutoff value before reaching an upper cutoff value where the ligand is considered unbound. The bound state in experiments is defined as the point where the labelled ligand becomes fluorescent. Since this point is not exactly known, the committor functions were calculated for multiple lower cutoff values. The ratio of the unbinding committor and the binding committor is shown in figure 5.13.

Can these committor functions be interpreted in terms of a free energy barrier as sketched in figure 5.1? We tested this by simulating a one-dimensional Brownian dynamics process, for which we measured how often a barrier was crossed. The barrier was chosen as a simple inverse harmonic potential $U(x) = -\frac{k}{2}(x - x_0)^2$ with two free parameters k and x_0. For the Brownian dynamics, the committor curves were calculated identically as in the case of the MD simulations. The best agreement with the committor curves from the MD simulations was obtained for $k = 18\,{}^{\mathrm{kJ}}/_{\mathrm{mol\cdot nm^2}}$ and $x_0 = 0.45\,\mathrm{nm}$ (figure 5.14).

The committor functions of the one-dimensional model agree remarkable well with the ones from the MD simulations, especially given the simple functional form of the potential.

This shows that the dynamics in the binding funnel can be modelled by a one-dimensional free-energy landscape and furthermore that a simple harmonic shaped barrier suffices.

Although it is not clear how the different ligand configurations in the binding funnel have to be mapped onto the one-dimensional landscape, this result is strong evidence for the existence of a free energy barrier within the binding funnel.

5.4.4.4. Combining Individual Terms for On-Rate Estimation

Putting all terms from equation (5.3), i.e., the effective target radius $a(x)$ and the translational part of the onrate, $k_{\mathrm{translational}} = 4\pi N_A D a(x)$, the rotational confinement $\frac{\Omega(x)}{4\pi}$ and the binding barrier crossing probability $p(\mathrm{bind} \mid RMSD = x)$ together, the free parameter x remains. Furthermore the probability of binding depends on the exact lower cutoff, which corresponds to the point in the experiments where a labelled ligand becomes fluorescent.

Does the choice of the point x influence the on-rate?

To answer this, we put all three terms together and calculated the overall on-rate in dependence of x and the lower cutoff (figure 5.15). There are several remarkable features:

First, for a large range the on-rate estimate does not depend strongly on the choice x of the ligand binding RMSD. The drop in the estimated on-rate for large RMSD close to $x = 1\,\mathrm{nm}$ stems from the fact that the "unbound state cutoff" was defined at 1 nm, i.e., that by construction $p(\mathrm{bind} \mid \mathrm{RMSD} = 1\,\mathrm{nm}) = 0$, thus also

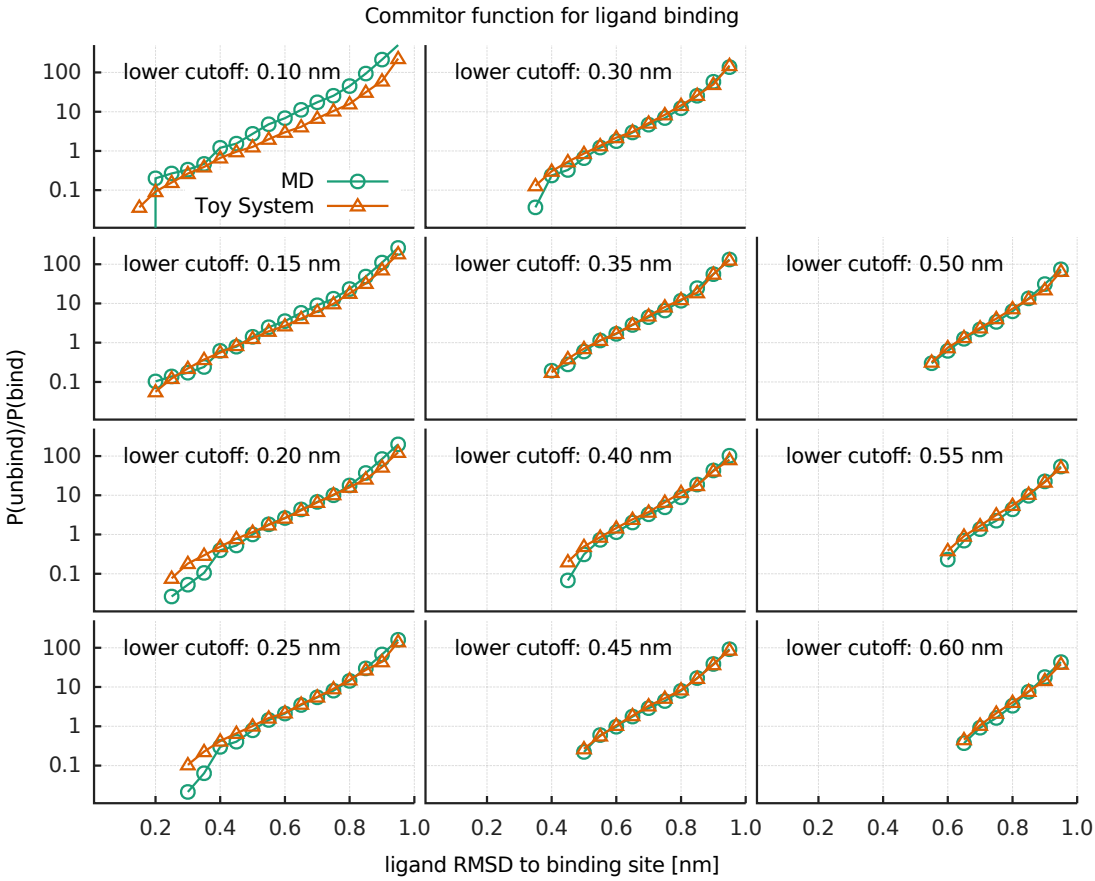


Figure 5.14.: Committor functions calculated identically to the ones in figure 5.13 for Brownian dynamics simulations in an inverse harmonic potential $U(x) = -\frac{k}{2}(x-x_0)^2$ yield similar unbinding/binding committor curves (orange) as MD simulations of ligand binding (turquoise). The parameters k and x_0 were determined by a fitting procedure. The high similarity shows that the ligand dynamics in the funnel is equivalent to one-dimensional Brownian dynamics and renders the assumption of a free energy barrier highly plausible.

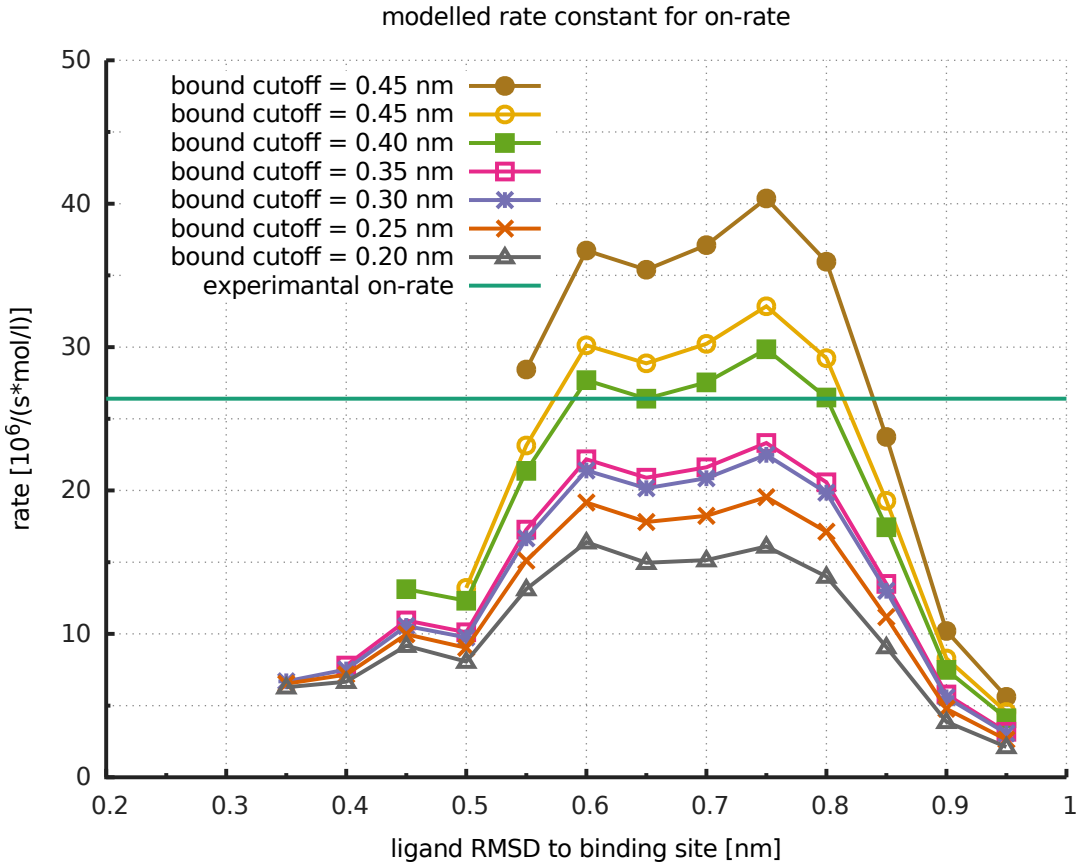


Figure 5.15.: Estimated effective on-rate calculated via equation 5.3 in dependence of the point x where the funnel model "switches" from "target sphere diffusion" to "barrier crossing". The calculated on-rate is determined for different cutoff values for the bound state. This cutoff corresponds to the point in which a labelled ligand would become fluorescent. The calculated on-rate depends within a large range only weakly on the switching point and varies within the range of reasonable bound state cutoff values only within a factor of 2.

$k_{\text{on}} = 0$ since $k_{\text{on}} \propto p(\text{bind})$. Second the variation in the estimated on-rate for different cutoff values is less than a factor of two. Third the predicted on-rate is similar to the experimental on-rate.

The first result shows that the model of a binding funnel is consistent. The observed barrier most likely stems from the conformational restriction. For small RMSD values ($x < 0.45\,\text{nm}$) the rate estimate drops slightly off, indicating that the approximation looses accuracy in this region. This is by no means surprising, as the Brownian dynamics simulations already suggested a transition state at $x = 0.45\,\text{nm}$. Past this transition state enthalpic effects become dominant and the free diffusion part of the model becomes increasingly unrealistic.

In the experimental setup the bound state is only defined by the fluorescence of the dye attached to the ligand. However, it is unknown in which exact conformation the dye is fluorescent. This seemingly impedes a direct comparison between simulation and experiment as the cutoff for the bound state is unknown. The second result however shows that within the range of reasonable lower cutoff values, the resulting on-rate varies only within a factor of 2.

The comparison of the MD-committor functions with the committor functions from the one-dimensional Brownian dynamics simulations suggest that the transition state is at 0.45 nm.

Third and most importantly the predicted on-rate, in the range between 15 and $30\,\mu\text{s}^{-1}(\text{mol/l})^{-1}$, is in very good agreement with the experimental on-rate of $26\,\mu\text{s}^{-1}(\text{mol/l})^{-1}$ (Peuker et al., 2013). Given the estimated experimental accuracy is also roughly within a factor of 2 this is an excellent support for the funnel hypothesis.

5.5. Conclusion

In this chapter we investigated the binding of cAMP at the CNBD of the MloK1 ion channel using MD simulations. We established a microscopic picture of ligand binding at high ligand concentrations. We then built a model describing the relevant mechanics of the process for low concentrations and predicted an on-rate for low ligand concentrations.

The first goal was to identify substates and transition probabilities between these substates. Using massive sampling and Markov modelling of the ligand dynamics, we found besides the ligand binding site several surface binding spots. For the high ligand concentration for which the simulations were carried out the timescale for transitions to the binding site were in the order of $2\,\mu\text{s}$, similarly the timescales for transitions between the surface binding spots was also in the range of up to $1\,\mu\text{s}$. The conclusion here is that for high ligand concentrations the surface plays a crucial role and the binding process can no longer be described as a simple two-state process.

For low ligand concentrations experiments had shown that the process is well described by a two-state process with a surprisingly low on-rate. To explain this

low rate we formulated and tested two models: a three-state model, where the low on-rate stems from slow transitions on the protein surface into the binding site and a funnel model, where rotational and translational confinement determine the on-rate.

We first found that the three-state model would predict an on-rate that is not compatible with the on-rate observed in experiments, ruling out this model.

In the alternative model the ligand binding is described by a binding funnel with the binding site at its bottom. The funnel increasingly confines the available conformational space of the ligand. The model combines a Smoluchowski type diffusional term with a free energy barrier crossing term. The diffusion term describes the diffusion into smaller and smaller volume elements in the binding funnel, combined with the increasing rotational restriction. The free energy term describes the probability to cross the remaining free energy barrier into the binding site.

Extracting all relevant parameters from the MD trajectories, we estimated the on-rate constant to be in an interval between 15 and 35 $\mu s^{-1}(mol/l)^{-1}$. This on-rate constant is in nice agreement with the experimentally determined value of $k_{on} = 26\,\mu s^{-1}(mol/l)^{-1}$.

The agreement with the measured on-rate is strong support for the existence of the binding funnel. The internal consistency shows that the assumption of a free energy barrier is justified and that the uphill part of the barrier is mainly entropic, i.e. caused by the translational and rotational confinement. Our comparison with Brownian dynamics simulations furthermore strongly suggest that the dynamics in the funnel is equivalent to the dynamics of overdamped diffusion over a one-dimensional barrier.

The results sketched here are of course specific to the binding of cAMP to the CNBD of MloK1. It is however highly likely that the same binding model, albeit with other parameters, is also valid for the binding of other secondary messengers such as cyclic guanosine monophosphate at the CNBD of MloK1. In how far these findings and the simple funnel model are transferable to other ligand-protein binding systems remains a topic for future investigation. The nature of the deeply buried binding pocket however motivates the speculation that the model is valid also in other cases of similar deep binding sites.

Part III.

Conformational Flexibility and Long-range Allostery of CRM1

6. Introduction

Eukaryotic cells are divided into separated compartments. Maybe the most important case is the nucleus that contains the majority of the cell's genetic material. The interior of the nucleus is separated from the cytoplasm by the nuclear envelope, a double lipid membrane. Exchange between nucleoplasm and cytoplasm is enabled by the nuclear pore complexes (NPCs). Small molecules up to a molecular mass of 30 kDa diffuse through the nuclear pore complex. Larger macromolecules, however, require the assistance of transport proteins (Cook and Conti, 2010).

Most transport proteins belong to the karyopherin-β superfamily (Görlich and Kutay, 1999; Cook et al., 2007). Depending on the directionality of the transport they are referred to as importins or exportins. A prototypic exportin is CRM1, also referred to as exportin-1 (Fornerod et al., 1997; Stade et al., 1997)

CRM1 transports proteins bearing a region called nuclear export signal (NES) by binding them in a specific binding site, the NES-cleft (Güttler et al., 2010; Dong et al., 2009; Monecke et al., 2009) and simultaneously binding the GTPase Ran in its GTP-bound form (RanGTP). The binding of cargo and RanGTP is cooperative, i.e., the affinity of the one binding partner is increased upon the binding of the other (Askjaer et al., 1998; Petosa et al., 2004).

This allosteric binding is key to the transport mechanism, sketched in figure 6.1: In the nucleus the RanGTP concentration is high, thus allowing binding of CRM1 and RanGTP. Due to the allostery this facilitates the formation of a ternary complex of CRM1, RanGTP and cargo which then can transverse the NPC. In the cytoplasm the RanGTP concentration is low, facilitating the disassembly of the complex and thus the release of the cargo.

In the past, several x-ray structures of CRM1 have been solved (Güttler et al., 2010; Dong et al., 2009; Monecke et al., 2009), including the ternary complex of CRM1, RanGTP and snurportin1 (SPN1) (Monecke et al., 2009). They revealed that CRM1, similar to other karyopherins, consists of subunits, so called HEAT repeats, which are built by two antiparallel α-helices and together form a toroidal, ring-like structure. It has been found that in complex with RanGTP and SPN1 CRM1 forms a compact conformation where RanGTP and SPN1 bind at spatially separated sites: While RanGTP binds in the middle of the of the ring, the cargo protein (in this case SPN1) binds on the outside between two HEAT repeats (figure 6.2 A).

The large distance between the two binding sites raises the major question: What is the underlying mechanism for the observed long-range allostery?. This question will form the centre of our MD study of CRM1.

An important experimental first step laid out in detail in the first publication is

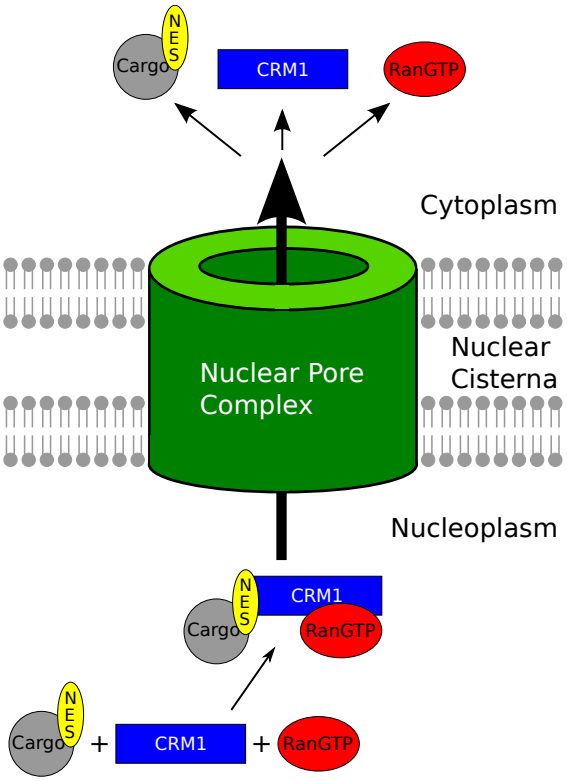


Figure 6.1.: Schematic sketch of the transport of cargo proteins through the nuclear pore complex via CRM1. In the nucleoplasm (bottom) the RanGTP concentration is high. As a consequence and via allostery CRM1, RanGTP and cargo form a complex which then can pass the nuclear pore complex. In the cytoplasm (top) the RanGTP concentration is low, shifting the equilibrium to the disassembled state.

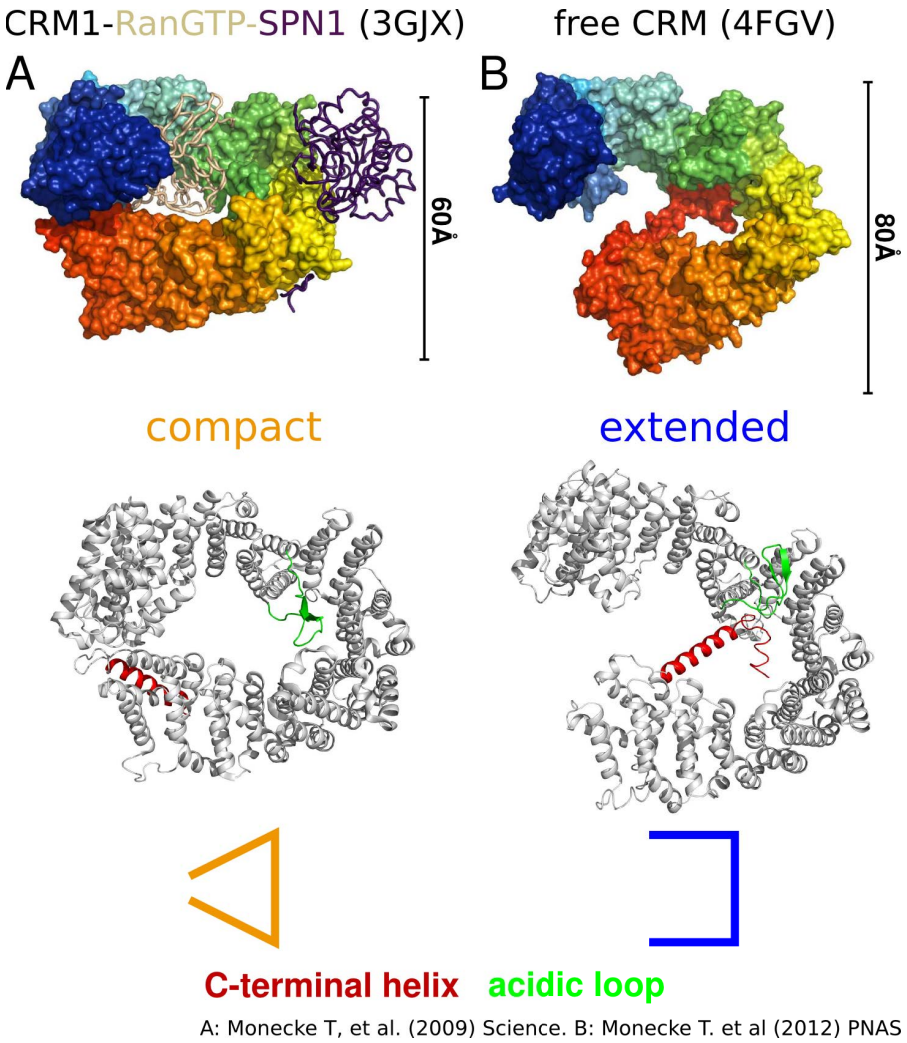


Figure 6.2.: X-ray structures of CRM1 in complex with SPN1 and RanGTP in the compact conformation (A) and free CRM1 in the extended conformation (B).

the determination of the structure of free unbound CRM1 via x-ray crystallography: Unlike in all other structures, free CRM1 adopts an extended, superhelical conformation (figure 6.2 B).

What is the reason for the different conformations? Two features of CRM1 differ markedly between the compact and extended conformation. The first is the arrangement of a long C-terminal helix which is parallel to the HEAT-repeat helices in the compact, bound conformation but crosses the ring in the free extended conformation. The second is the conformation of a long, acidic loop that is "flipped-back" in the extended conformation and crosses the ring like a seat belt in the compact conformation of the ternary complex.

Motivated by the obvious difference in the two conformations, we postulate the hypothesis that one or both of these features are linked to the overall conformational change.

Another feature of the x-ray structures is the NES-cleft, i.e. the cargo binding site, which assumes an open state in the compact, cargo bound conformation and a closed state in the extended, free structure. This leads to our second hypothesis, namely there is a dependence between the overall conformation and the NES-cleft conformation.

To test these two hypotheses, we investigated the effect of C-terminal helix and acidic loop on the overall conformation via MD simulations of in-silico mutants and we used MD simulations to study the dependence of the binding site conformation on the overall conformation.

These test are addressed in the first publication.

The overall conformation turns out to influence the free energy differences between the cargo binding site states. However, no direct influence of the acidic loop on the extended conformation is found. Beside its differing conformation though, the acidic loop has been shown to have dramatic influence on the cooperativity of cargo and RanGTP binding (Petosa et al., 2004). To understand why this is the case, two hypothesis are formulated and tested: Either the acidic loops arrangement in the compact CRM1 conformation stabilises this conformation, or it transfers the information of the overall conformation to the binding site. Both possibilities are investigated using MD simulations to elucidate the role of the acidic loop and thus complete the picture of the allosteric binding.

These investigations are part of the second publication.

6.1. Own Contribution

The following two publications are the result of a collaboration between researchers from different groups. My own contribution focused on the question of the underlying mechanism of the long-range allostery, especially on the role of the C-terminal helix and the acidic loop as well as the determination of the structural stability of free CRM1.

To address these issues, I prepared, carried out, analysed, interpreted all MD

simulation parts and wrote and discussed the corresponding parts of the PNAS publication (Monecke et al., 2013).

In the second publication (Dölker et al., 2013) I prepared, carried out, analysed, interpreted all simulations analysing the influence of the acidic loop (and RanGTP) on the NES-cleft configuration, as well as the force probing simulations of the acidic loop deletion mutant of compact conformation CRM1 and the analogue comparison simulations of wild type CRM1 and wrote and discussed the corresponding sections in the paper. These results are discussed in the subsection *MD Analysis of Structural Changes in Ran and NES Binding Sites*. I furthermore contributed to the writing of the results, discussion and conclusion sections of the other MD simulation parts.

Structural basis for cooperativity of CRM1 export complex formation

Thomas Monecke[a], David Haselbach[b], Béla Voß[c], Andreas Russek[c], Piotr Neumann[a], Emma Thomson[d], Ed Hurt[d], Ulrich Zachariae[e], Holger Stark[b,f], Helmut Grubmüller[c], Achim Dickmanns[a,1], and Ralf Ficner[a]

[a]Abteilung für Molekulare Strukturbiologie, Institut für Mikrobiologie und Genetik, Göttinger Zentrum für Molekulare Biowissenschaften, Georg-August-Universität Göttingen, D-37077 Göttingen, Germany; [b]Dreidimensionale Kryo-Elektronenmikroskopie, Max-Planck-Institut für Biophysikalische Chemie, D-37077 Göttingen, Germany; [c]Abteilung für Theoretische und Computergestützte Biophysik, Max-Planck-Institut für Biophysikalische Chemie, D-37077 Göttingen, Germany; [d]Biochemie-Zentrum der Universität Heidelberg, D-69120 Heidelberg, Germany; [e]Scottish Universities' Physics Alliance, School of Physics and Astronomy, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom; and [f]Abteilung für Molekulare Kryo-Elektronenmikroskopie, Institut für Mikrobiologie und Genetik, Göttinger Zentrum für Molekulare Biowissenschaften, Georg-August-Universität Göttingen, D-37077 Göttingen, Germany

Edited by Robert Huber, Max Planck Institute of Biochemistry, Planegg-Martinsried, Germany, and approved December 3, 2012 (received for review September 3, 2012)

In eukaryotes, the nucleocytoplasmic transport of macromolecules is mainly mediated by soluble nuclear transport receptors of the karyopherin-β superfamily termed importins and exportins. The highly versatile exportin chromosome region maintenance 1 (CRM1) is essential for nuclear depletion of numerous structurally and functionally unrelated protein and ribonucleoprotein cargoes. CRM1 has been shown to adopt a toroidal structure in several functional transport complexes and was thought to maintain this conformation throughout the entire nucleocytoplasmic transport cycle. We solved crystal structures of free CRM1 from the thermophilic eukaryote *Chaetomium thermophilum*. Surprisingly, unbound CRM1 exhibits an overall extended and pitched superhelical conformation. The two regulatory regions, namely the acidic loop and the C-terminal α-helix, are dramatically repositioned in free CRM1 in comparison with the ternary CRM1–Ran–Snurportin1 export complex. Single-particle EM analysis demonstrates that, in a noncrystalline environment, free CRM1 exists in equilibrium between extended, superhelical and compact, ring-like conformations. Molecular dynamics simulations show that the C-terminal helix plays an important role in regulating the transition from an extended to a compact conformation and reveal how the binding site for nuclear export signals of cargoes is modulated by different CRM1 conformations. Combining these results, we propose a model for the cooperativity of CRM1 export complex assembly involving the long-range allosteric communication between the distant binding sites of GTP-bound Ran and cargo.

Eukaryotic cells are separated into compartments, such as the endoplasmic reticulum, mitochondria, or the nucleus. The nucleus is encompassed by a double-layered membrane, the nuclear envelope, whereas aqueous connections to the cytoplasm are maintained by large macromolecular assemblies, the nuclear pore complexes. However, aside from the regulatory advantages of compartmentalization, the spatial separation also poses a logistic challenge, namely to ensure the efficient exchange of proteins, RNA, and metabolites between these compartments. An elaborate transport system has evolved to achieve the bidirectional transport of proteins and RNAs. Nuclear transport of macromolecules in eukaryotic cells is mainly mediated by soluble transport receptors of the karyopherin-β superfamily termed importins and exportins (1, 2). They share a common structural arrangement of approximately 20 repetitive elements, so-called HEAT repeats. Among them, the prototypical exportin chromosome region maintenance 1 (CRM1) is essential for nuclear depletion of numerous structurally and functionally unrelated protein and ribonucleoprotein cargoes (3–7). Commonly, CRM1–cargo binding depends on a 10- to 15-residue-long, leucine-rich nuclear export signal (NES) within the transport target (8–10), which binds to a hydrophobic cleft (NES binding cleft) on the outer convex surface of CRM1. Efficient cargo binding requires the presence of the small GTPase Ran in its GTP-bound nuclear form (RanGTP). The binding of RanGTP and cargo to CRM1 has been shown to be cooperative, as the affinity for either of the binding partners is increased in the presence of the other protein (11–13). However, the structural basis for this cooperativity has so far been unclear. Subsequent to formation of the stable CRM1–RanGTP–cargo complex, the assembly translocates through the nuclear pore complex into the cytoplasm. GTP hydrolysis by Ran in the cytoplasm is stimulated by the Ran GTPase activating protein (i.e., RanGAP) and further increased by Ran binding proteins (RanBPs), resulting in a release of the cargo. CRM1 in the free form shuttles back into the nucleus for the next round of export.

In recent years, the crystal structures of three different CRM1 export complexes have been solved. They are CRM1 with bound cargo Snurportin1 [CRM1–SPN1; Protein Data Bank (PDB) ID code 3GB8] (10), CRM1 with RanGTP (CRM1–RanGTP; PDB ID code 3NC1) (8)—both representing assembly intermediates—as well as the functional ternary export complex CRM1–RanGTP–SPN1 (PDB ID code 3GJX) (9). Moreover, the structure of one disassembly complex containing CRM1, RanGTP, and RanBP1 (CRM1–RanGTP–RanBP1; PDB ID code 3M1I) (14) has been determined. All these structures have in common that CRM1 adopts a compact ring-like shape of a toroid with the N- and C-terminal regions forming numerous interactions. Alterations between these CRM1 structures are found in the first three HEAT repeats, a highly conserved region involved in RanGTP binding, but the N- and C-terminal HEAT repeats are always in close contact. This led to the proposal that cargo-free CRM1 may retain a ring-like shape (9), which was further supported by single-particle EM and small-angle X-ray scattering analyses (12, 15).

Surprisingly, and in contrast to all other known export complexes, no obvious explanation for the observed cooperative effects were seen; in particular, no direct interactions between cargo and RanGTP in the CRM1–RanGTP–SPN1 complex could be detected. Therefore, two significant differences between these ring-shaped CRM1 structures might be of particular relevance. The first one concerns the so-called acidic loop, a stretch of acidic amino acids forming a more or less extended loop of variable length in many members of the importin-β

Author contributions: T.M., P.N., U.Z., H.S., H.G., A.D., and R.F. designed research; T.M., D.H., B.V., A.R., P.N., and A.D. performed research; E.T. and E.H. contributed new reagents/analytic tools; T.M., D.H., B.V., A.R., P.N., U.Z., H.S., H.G., A.D., and R.F. analyzed data; and T.M., D.H., B.V., P.N., U.Z., H.S., H.G., A.D., and R.F. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 4FGV and 4HZK); and the EMDataBank (ID codes EMD-2110 and EMD-2111).

[1]To whom correspondence should be addressed. E-mail: adickma@uni-goettingen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1215214110/-/DCSupplemental.

superfamily, which has been shown to be involved in cargo binding and release (9, 16–18). In the RanGTP-bound CRM1 complex structures, this acidic loop traverses the central opening of the CRM1 toroid and affixes RanGTP to CRM1, in a seatbelt-like fashion, to N- and C-terminal regions that form numerous interactions (8, 9). In the CRM1–RanGTP–RanBP1 complex, this loop is reoriented in a "flipped back" position that brings the loop into proximity of the HEAT repeats, forming the NES binding cleft (14). The binding of the acidic loop to the inward-oriented HEAT repeat helices 11B and 12B may stabilize the NES cleft in a contracted, closed state. The other feature in CRM1 relates to the last HEAT repeat helix 21B, which is located in a parallel orientation to the HEAT repeat helix 21A in the Ran-bound complexes (9, 14). In the cargo-only bound state (i.e., CRM1–SPN1), these two helices are considerably repositioned, and the relative position of the N-terminal HEAT repeats with respect to the C-terminal region and thus the interaction pattern of these two regions is altered (10). The B-helix is now bridging the central opening and interacts with the HEAT repeats that form the NES cleft. The acidic residues located C-terminally of the B-helix (i.e., C-terminal acidic tail) are thought to form close contacts to basic patches at the back side of the NES binding cleft, thereby regulating the NES binding cleft state (19). However, the question of how CRM1 achieves these conformational changes and how binding cooperativity is mediated from the RanGTP binding site to the cargo binding site remains elusive, mainly because no crystal structure of free CRM1 has yet been solved.

Here, we present crystal structures of free CRM1 exhibiting an extended and pitched superhelical conformation. Single-particle EM studies reveal that, in a noncrystalline environment, free CRM1 exists in equilibrium between extended and more compact conformations. Moreover, molecular dynamics (MD) simulations unravel the contributions of two regulatory regions to the conformational state of CRM1. Combining these results, we propose a model for the cooperativity of CRM1 export complex assembly, which shows how long-range allosteric communication between distant binding sites in proteins orchestrates the tuning of affinities.

Results

Crystal Structure Analysis. To gain insight into the structure of free CRM1, and to understand the structural requirements for the cooperativity of cargo and Ran binding, we aimed to crystallize it. As all attempts to crystallize full-length mammalian CRM1 have failed so far, we sought to structurally characterize the CRM1 orthologue from the thermophilic fungus *Chaetomium thermophilum* (*ct*CRM1) which exhibits a sequence identity of 50% to human CRM1 (Fig. S1). Two different crystal forms of *ct*CRM1 were obtained belonging to space groups $P3_1$ and $P2_12_12_1$, respectively. As a result of better crystal quality and higher resolution of the diffraction data, as well as the fact that the trigonal crystals were twinned, the orthorhombic crystals were initially used. The structure was solved by molecular replacement using mouse CRM1 derived from the CRM1–RanGTP–SPN1 complex structure (9) as search model, and refined at 2.94 Å resolution to R and R_{free} values of 22.0% and 24.3%, respectively (Table S1).

Like mammalian and yeast CRM1, full-length *ct*CRM1 consists of 21 HEAT repeats (H1–H21; Fig. S1). It adopts an overall extended and pitched superhelical structure (Fig. 1), clearly differing from the known complex structures, in which CRM1 is bent into a distorted toroid and shows less superhelical pitch (9, 10, 14) (Fig. 2 *B–D*). In all these complexes, N- and C-terminal HEAT repeats of CRM1 interact with each other, involving residues of H2 to H5 on the one side and residues from H21 on the other side. These interactions are completely absent from the cargo-free form of CRM1 (Figs. 1 and 2*A*). In contrast to the complex structures, the C-terminal half of free CRM1 is shifted from the N-terminal HEAT repeats by approximately 20 Å, increasing the pitch and elongating the superhelix by 12 Å (Fig. 2). As a consequence, the helices 11A and 12A move toward each other, thereby rendering the NES binding cleft between these helices inaccessible for cargo (Fig. 3 and Fig. S2).

The extended conformation of free CRM1, and, consequently, the closed state of the NES binding cleft, seems to be stabilized by the rearranged C-terminal helix of CRM1 (HEAT repeat helix 21B), which diagonally crosses the superhelix. The C-terminal helix is found in a similar position as in the CRM1–SPN1 binary complex (10) (Fig. 1 and Fig. S3). Several residues between Arg1059 and Glu1073 belonging to the C-terminal helix and the adjacent C-terminal tail contact the B-helices of H8 to H12 on the opposite side of CRM1 (Fig. 1). Only one salt bridge each has been observed for the C-terminal tail in interaction with the acidic loop (Lys1068 with Glu426) and the back side of the NES cleft (Glu1073 and Gln591; Fig. 1). In contrast, the B-helix of H21 is located at the outer surface of the CRM1 toroid in parallel orientation to helix 21A in the Ran-containing complexes (Fig. S3).

Another important structural difference between free CRM1 and its export complexes concerns the acidic loop, which is inserted between the A- and B-helix of H9 (Figs. S1 and S4). In free *ct*CRM1, this region contacts the HEAT repeat helices forming the back side of the NES cleft (Fig. S5*A*). The observed

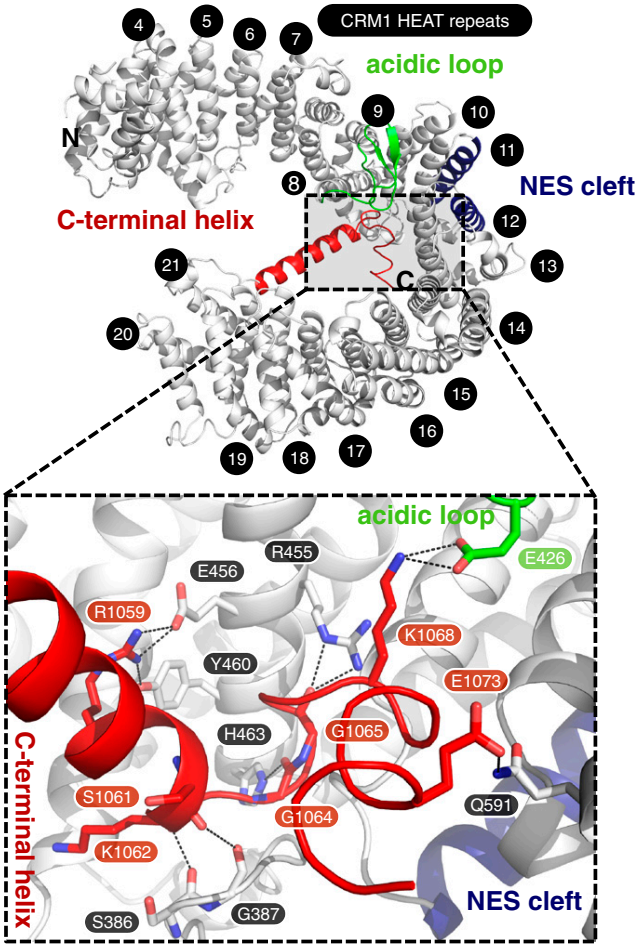


Fig. 1. Crystal structure of free *ct*CRM1 (gray). The acidic loop (green), the C-terminal helix (red) and the NES cleft (blue) are highlighted. The HEAT repeats are numbered, and termini are labeled. (*Lower*) Rotated detail view of the interactions of residues from the C-terminal helix with the acidic loop and a patch of CRM1 formed by helices of H8 to H12. Hydrogen bonds and salt bridges are represented by dashed lines, and interacting residues are labeled.

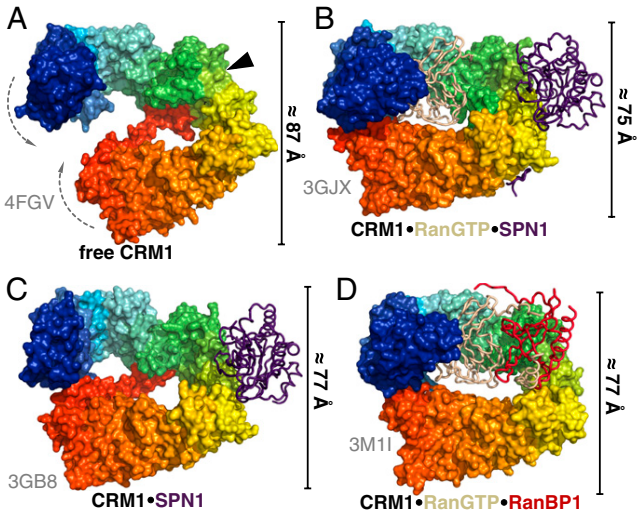


Fig. 2. Comparison of CRM1 conformations in different crystal structures. Free *ct*CRM1 (*A*), CRM1–RanGTP–SPN1 (*B*), CRM1–SPN1 (*C*), and CRM1–RanGTP–RanBP1 (*D*) are shown. CRM1 is depicted as rainbow colored surface from N (blue) to C terminus (red), whereas the interacting proteins are shown as tube models (SPN1, purple; RanGTP, beige; RanBP1, red). The position of the NES cleft in structures lacking cargo is marked by a black arrowhead. The bars at the side of the individual structures indicate the dimension of the protein in the shown orientation.

conformation of the acidic loop is clearly different from that in the CRM1–RanGTP and CRM1–RanGTP–SPN1 complexes (in the seatbelt conformation) and instead closely resembles the orientation of the respective part in the CRM1–RanGTP–RanBP1 structure, with the acidic loop in a flipped-back conformation (Fig. S3). Specifically, hydrophobic residues of the acidic loop (Val427, Leu428, Ile429, and Ile437) pack against the surface of helices 10B, 11B, and 12B at the back side of the NES cleft. This results in a major rearrangement of the side chains located in the hydrophobic core of H10, H11, and H12 (Fig. S5 *B–D*). Met580 is shifted toward the hydrophobic residues of the acidic loop, and the resulting empty space is filled by Met542, the position of which in turn is replaced by Phe569. As a consequence, this rearrangement facilitates a significant rotation and rearrangement of the highly conserved Lys531 and Lys534. The latter enables the helix 11A to move toward helix 12A. This ultimately leads to a closure of the NES cleft, which prevents binding of an NES-bearing cargo (Fig. S2). Mutations of the respective hydrophobic residues within the acidic loop of yeast CRM1 have been shown to cause a reduction of NES-cargo release rate in CRM1–RanGTP–RanBP1 complexes, indicating a direct influence on NES cleft accessibility (14).

However, analysis of the crystal packing of the orthorhombic crystal form revealed that this conformation of the acidic loop might also be stabilized by the interaction of Glu434, located in the acidic loop, with Gly642, of a symmetry-related molecule. To exclude the possibility that the flipped-back conformation of the acidic loop as well as the overall open conformation of CRM1 are at least in part products of crystal packing, the trigonal *ct*CRM1 crystal form was investigated as well. The 3.1-Å crystal structure of $P3_1$ *ct*CRM1 was solved by molecular replacement, revealing a different crystal packing of the CRM1 molecules (Table S1). There are two molecules in the asymmetric unit, both showing the extended superhelical structure with the C-terminal helix crossing the CRM1 arch and the acidic loop in the flipped-back position even though the acidic loop is not involved in any crystal contacts (Figs. S4*B* and S6*A*). However, there are also two remarkable differences between the structures. First, in the trigonal crystal form, CRM1 adopts a slightly less extended conformation with a reduced superhelical pitch, which also differs between the two molecules in the asymmetric unit, with an rmsd of 1.03 Å (Fig. S6*A*). Second, H1 to H3, as well as several loops and the C-terminal five residues corresponding to the C-terminal acidic tail, are not clearly defined in the electron density map and thus were not modeled. Overall, these differences reflect the intrinsic high plasticity of free CRM1, which is also known for other transport factors in their free state (20, 21).

EM Structure Analysis. We next asked whether the extended conformation of free CRM1 could also exist in a noncrystalline environment. For this purpose, *ct*CRM1 was subjected to the GraFix approach (22) and subsequent single-particle EM analysis (Figs. S7 and S8). Strikingly, free *ct*CRM1 was detected in at least two different and clearly distinct conformations. Two thirds of the classified particles clearly adopt an extended and pitched superhelical shape similar to that seen in the crystal structures of free *ct*CRM1 when fitted in the EM model (Fig. 4). Interestingly, the EM model supports a free CRM1 conformation, which adopts an even more extended state than seen in the orthorhombic crystal structure. The other conformer, represented by the remaining third, resembles the shape of a more compact closed ring or distorted toroid, reminiscent of the CRM1 conformation observed in various binary and ternary complexes. These data suggest that free CRM1, also in solution, is able to switch between an extended and a compact conformation.

MD Simulations. We next addressed the question whether CRM1 in the extended form represents a strained conformation, and, if so, which structural features prevent closure. We applied MD simulations, which have previously provided insight into the high conformational flexibility of HEAT repeat proteins (23–25). We first compared simulations of WT CRM1 with a deletion mutant lacking the C-terminal helix 21B and the acidic tail. To monitor transitions from the extended toward the compact crystal structure, the progress of the trajectory along the difference vector between those two structures was recorded and used as a reaction coordinate. In five unperturbed simulations of WT, all structures remain in the extended conformation or elongate even further (Fig. 5*A*). By contrast, 6 of 10 simulations of the helix deletion mutant spontaneously progress toward the compact state within as little as 100 ns (Fig. 5*B*). Simulations, in which closing and subsequent contact formation between the N- and C-

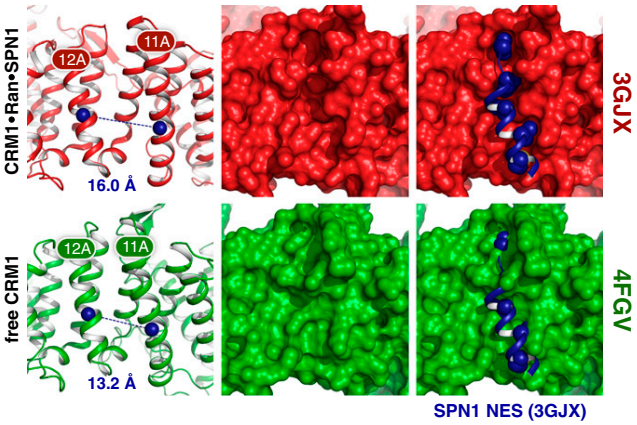


Fig. 3. Comparison of NES cleft conformations between free *ct*CRM1 (PDB ID code 4FGV; green) and CRM1 bound to SPN1 and RanGTP (PDB ID code 3GJX; red). The NES cleft is shown in cartoon mode (*Left*) with the centers of mass of the helices 11A and 12A represented by blue spheres and their distances indicated. A surface model (*Center*) illustrates the differences of the NES clefts between both structures. A superposition of the SPN1 NES from the ternary CRM1–RanGTP–SPN1 complex (*Right*) highlights the structural changes in the NES cleft, which are incompatible with NES binding in free CRM1.

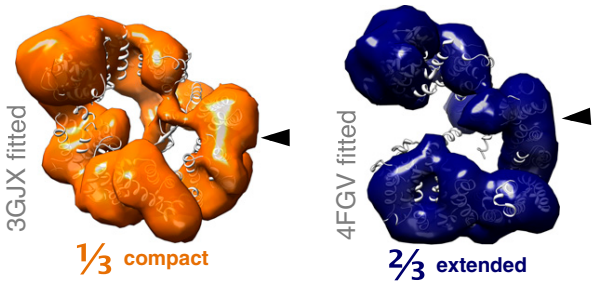


Fig. 4. Single-particle EM analysis of free *ct*CRM1. EM models of the compact (orange) as well as the extended conformation (blue) of free *ct*CRM1 are shown. The crystal structures of free *ct*CRM1 and CRM1 in complex with SPN1 and RanGTP are fitted to the envelope models of the EM structures. The position of the NES cleft is marked by a black arrowhead.

terminal regions of CRM1 are observed, display lower flexibility after closing than those that remain open. These results show that the C-terminal helix is a major determinant restricting the conformational flexibility and thus shifting the population more toward the extended conformation.

We next examined the contribution of the acidic loop. In all but one simulation of the acidic loop deletion mutant, the structure remains in the extended conformation, closely resembling the behavior of WT (Fig. S9*A*). In contrast, the double deletion mutant lacking both the C-terminal helix and the acidic loop shows conformational changes similar to the deletion mutant lacking the C-terminal helix only (Fig. S9*B*). This result indicates that the acidic loop plays only a minor role in determining the conformation of CRM1. To characterize the closing motion leading from the extended to compact conformation in more detail, all trajectories were projected onto the subspace defined by the three available crystal structures of free CRM1, the CRM1–SPN1 complex, and the ternary CRM1–RanGTP–SPN1 complex (Fig. 5*C*). Starting from the extended state, CRM1 does not seem to target the compact conformation seen in the ternary CRM1–RanGTP–SPN1 complex, but rather approaches an almost compact conformation resembling the one seen in the CRM1–SPN1 complex (Fig. 2*C*). Indeed, the rmsd to the almost compact crystal structure decreases from an initial value of 5.1 Å to a minimum of 2 Å, whereas that to the compact conformation remains at a minimum of 3.9 Å.

Because the plane defined by the three aforementioned crystal structures may not contain all relevant dynamics, and to identify the dominant structural changes, we performed a principal component analysis of the combined ensemble of all trajectories of the C-terminal deletion mutant. Fig. S9*C* shows projections onto the first two eigenvectors of all trajectories, including all variants. As indicated by the inserted ribbon structures, motions along the first principal component analysis eigenvector predominantly describe a change in the diameter of the ring, whereas dynamics along the second eigenvector represent a change in the pitch of the superhelical structure (Movies S1 and S2). Whereas the first eigenvector is similar to the vector connecting the extended and compact crystal structures, significant dynamics are also seen along the second eigenvector, i.e., changes in the pitch without changes in the ring diameter. This motion is not contained in the differences among the three crystal structures.

Next we asked whether the conformational transition from the almost compact to the extended state of CRM1 may affect the cargo affinity of the CRM1 NES binding cleft. To this end, we examined structural changes of the binding cleft and investigated if these depend on the CRM1 conformation. Accordingly, each of the trajectories was subdivided into four states—extended/open, extended/closed, compact/open, and compact/closed—corresponding to the respective state of the overall CRM1 conformation and that of the NES binding cleft, respectively. We then determined from the simulations the conditional probabilities that the binding cleft is open (i.e., high-affinity state), given that the overall conformation is extended or, respectively, compact. Indeed, the probability of observing an NES binding configuration of the NES cleft is larger by a factor of approximately three for the compact CRM1 conformation than for the extended, low-affinity one, with significances lower than 1% and 3% for the double deletion and helix deletion mutant, respectively (Table S2). This population shift corresponds to a marked increase of the free energy difference between the closed and open NES cleft conformations of $\Delta\Delta G$ of 2.7 ± 1.2 kJ/mol for the C-terminal helix deletion mutant and of $\Delta\Delta G$ of 3.5 ± 1.3 kJ/mol for the double deletion mutant.

Discussion

The structural basis for the cooperative binding of cargo and RanGTP to the exportin CRM1 has so far remained an open question. All crystal structures of CRM1 in various complexes with RanGTP and/or SPN1 or RanGTP and RanBP1 showed a toroid, a distorted ring-shaped molecule, in which N- and C-terminal HEAT repeats interact (Fig. 2 *B–D* and Fig. S3).

Here we demonstrate that CRM1 in the free state can adopt a compact, ring-like structure, and an extended, superhelical structure. The structures of free CRM1 obtained from different crystal forms, as well as single-particle EM and MD simulations, indicate a high intrinsic flexibility. The MD simulations show that the C-terminal helix contributes to stabilizing the extended

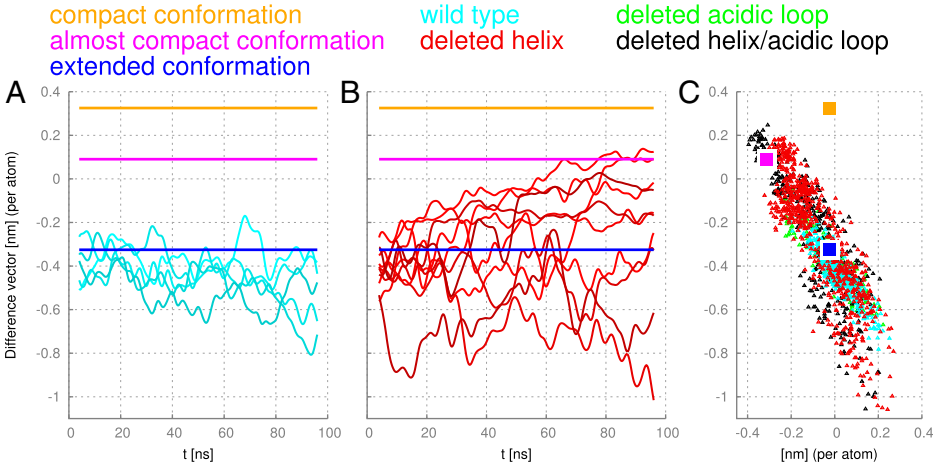


Fig. 5. MD simulations of WT and mutant free *ct*CRM1. Projections of WT simulations (*A*, cyan) and simulations with deleted C-terminal helix (*B*, red) onto the difference vector between the extended and compact structure constitute a measure of how much the protein changes into the compact conformation. (*C*) Projections onto the plane in the configurational space spanned by the extended, compact, and almost compact crystal structure show that, after deletion of the C-terminal helix, the system adopts the configuration of the almost compact structure (magenta square) rather than the compact conformation (orange square).

conformation like a ratchet, whereas the acidic loop is of less importance for the overall structure. Moreover, the extended conformation exhibits a high degree of flexibility that is lost in the closing process as a result of interactions between the terminal HEAT repeats that increase the rigidity of CRM1.

The cooperative binding of RanGTP and cargo by CRM1 was previously suggested to be achieved mainly by local subtle structural changes in the CRM1 toroid (9, 19). Binding of either of the two interaction partners leads to rearrangements in CRM1 altering the binding properties of the second binding site. However, in light of the results presented here, the cooperativity of cargo and RanGTP binding can occur in both pathways for export complex assembly that can be envisioned. As suggested by our EM data, free CRM1 is predominantly in an extended conformation, most likely with the acidic loop flipped back, intricately interacting with the HEAT repeats involved in NES binding, and stabilized by the C-terminal helix 21B (Fig. 6). For binding of RanGTP first, helix 21B has to be released and move into a parallel orientation to helix 21A, as the crossing C-terminal helix is partially occupying the volume required for RanGTP binding (Fig. S3). The release and reorientation of helix 21B seems to be an unfavorable process, which could be causative for the low micromolar affinity of RanGTP to CRM1 (11, 12). During RanGTP binding, the extended structure condenses by interaction of N- and C-terminal regions, both contacting RanGTP. Additionally, the acidic loop is released from the flipped-back position and moves toward the seatbelt position. As a consequence, the NES cleft retains its flexibility and becomes more accessible to cargo by changing the conformation, thereby enhancing the affinities for cargo by a factor of approximately 500 from a low micromolar to a low nanomolar range (11). These large structural rearrangements of the acidic loop and the C-terminal helix of H21 lead to a compact structure with only little superhelical pitch.

In the vice-versa assembly procedure, SPN1 binds first to CRM1 (Fig. 6). In general, NES binding is weak (i.e., micromolar K_d) in the absence of RanGTP, but might be increased as other domains of the cargo bind to CRM1 in addition (10, 11). The association of CRM1 and cargo requires CRM1 in a more toroid-like shape to open the NES binding cleft. The MD simulations suggest a pronounced population shift of the NES cleft toward its open, high-affinity conformation upon CRM1 compaction, which points to a conformational selection mechanism rather than a clear-cut induced fit. In this toroid conformation, the helix 21B might still span the CRM1 interior as observed in the CRM1–SPN1 binary complex structure (10), but is incompatible with the arrangement of the molecules as present in the ternary export complex of CRM1–RanGTP–SPN1 (Fig. S6*B*). The structure of CRM1 in this binary CRM1–SPN1 complex reveals a high flexibility of the terminal regions as well as the acidic loop, suggesting its release from the flipped-back position upon cargo binding. This conformation with the helix 21B in a kinked and presumably more strained conformation would increase the affinity for RanGTP ~500-fold (11, 12) by facilitating the release and rearrangement of helix 21B (Fig. S6*B*). Upon RanGTP binding, CRM1 has to arrest properly in the ring-like shape for a stable ternary complex formation.

Taken together, in both scenarios, upon binding of one protein, CRM1 is transformed from a more or less extended conformation to an almost compact one. This intermediate state with reoriented helix 21B and acidic loop changes the structural properties of the binding site for the missing partner and leads to an increase of affinity. In any case, the C-terminal helix plays a critical role as its deletion causes a dramatic decrease of cooperativity in complex formation and thus leads to high-affinity cargo binding even in the absence of RanGTP (26). Detailed biochemical analysis revealed that deletion of the last 9 aa of CRM1 (C-terminal acidic tail) or mutation of the acidic residues therein to alanine (^{1064}EIPEEMCD^{1071} to ^{1064}AIPAAMCA^{1071}) leads to an approximately 10-fold increase of CRM1-cargo affinity (15.6 μM to 1.7 μM), independent of the presence of

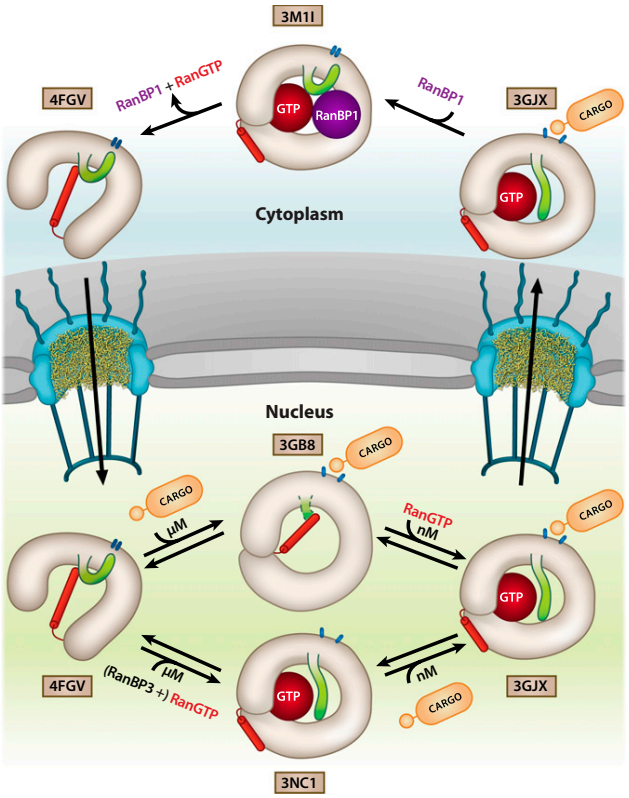


Fig. 6. Model for cooperative CRM1 export complex assembly and disassembly showing its conformational variability and the important structural features in different states of the transport cycle. CRM1 is shown in the respective conformations and colored in gray with the acidic loop highlighted in green. The C-terminal helix of CRM1 is shown in red, and the NES binding cleft is represented by blue ovals. The PDB ID codes of the individual crystal structures used are indicated.

RanGTP (19). Furthermore, this effect is dramatically increased when deletion of the C-terminal acidic tail is combined with mutations in the acidic loop (^{430}VLV^{432} to ^{430}AAA^{432}). In this double mutant, cargo affinity of free CRM1 is more than 600-fold higher with respect to WT (15.6 μM to 0.0025 μM), and thus comparable to the affinity of CRM1 for cargo in the presence of RanGTP (0.0015 μM for SPN1). However, of the acidic tail residues analyzed in this study, only Glu1073 of free *ct*CRM1 is in direct contact with the side chain of Gln591, which is the counterpart of Lys594 in human CRM1 (Fig. 1). This suggests a more global electrostatic attraction effect of positive and negative charged patches in the two regions of CRM1 rather than direct ionic interactions.

In summary, CRM1 exhibits large overall structural dynamics, in line with other transport receptors such as importin-β, exportin-t, and Cse1 (27). The cooperativity of cargo and RanGTP binding is achieved by the conformational switch of two distant regions, the acidic loop and the extensive motion of the C-terminal helix. This helix is a unique feature of CRM1 and represents an additional regulatory mechanism not observed in any other nuclear transport receptor. CRM1 dynamics reveal how the interplay between the distant and not-overlapping binding sites of RanGTP and cargo on CRM1 can be mediated and fine-tuned. It remains an open question whether the observed structural flexibility of CRM1 is also important for its function in cell cycle control during mitotic progression.

Materials and Methods

Detailed experimental procedures are described in *SI Materials and Methods*.

Protein Expression and Purification. GST-*ct*CRM1 was expressed from pET24d in *Escherichia coli* BL21(DE3) (Merck) at 20 °C. The protein was purified on a GSTrap column (GE Healthcare) followed by tobacco etch virus protease-mediated GST cleavage and a final gel filtration with a Superdex 200 column (GE Healthcare). Pure *ct*CRM1 was concentrated to 20 mg/mL and stored at −80 °C.

Crystallization and Crystal Structure Determination. *ct*CRM1 was crystallized by vapor diffusion. Orthorhombic crystals were flash-cooled in liquid nitrogen after soaking in reservoir solution containing, additionally, 14% (vol/vol) glycerol. A data set was collected at BL14.1 operated by the Helmholtz-Zentrum Berlin at the BESSY II electron storage ring (Berlin-Adlershof, Germany) (28). Data were processed with iMOSFLM (29) and SCALA (30). The structure was solved with PHASER (31) by using the crystal structure of mouse CRM1 as search model (PDB ID code 3GJX) (9). The model was refined with PHENIX (32) at 2.94 Å resolution to R and R_{free} values of 22.0% and 24.2%, respectively (Table S1). Data of the trigonal crystals were collected at beamline P14 (PETRA III, European Molecular Biology Laboratory, Hamburg, Germany) and processed by using XDS (33) and XSCALE. The structure was solved by means of molecular replacement with PHASER by using the previously solved orthorhombic *ct*CRM1 structure.

EM Preparation and Image Processing. *ct*CRM1 was subjected to the GraFix protocol (22). Negatively stained particles were imaged by using a CM200FEG (Philips) at a magnification of 155,000× (1.85 Å/pixel). Particles were picked, contrast transfer function correction was performed per Sander et al. (34), and further image processing was done in Imagic (35). A starting model was generated by using angular reconstitution facilitated by a voting algorithm (36). Two distinct conformations were detected in the dataset and it was split based on cross correlations. Final 3D models were obtained at a resolution of ∼20 Å.

MD Simulations and Analysis. Simulations were performed by using Gromacs 4.5 (37) together with the Amber99sb force field (38) and the extended simple point charge water model (39) in a constant particle number/pressure/temperature ensemble. Electrostatic interactions were calculated by using particle-mesh Ewald summation (40). We used four simulation systems: the WT CRM1 extended state (PDB ID code 4FGV), a C-terminal deletion mutant, an acidic loop deletion mutant, and a system containing both deletions. To compare our trajectories with the crystal structures available for CRM1, i.e., the CRM1–RanGTP–SPN1 complex (compact structure, PDB ID code 3GJX, from mouse), the CRM1–SPN1 complex (almost compact structure, PDB ID code 3GB8, from human), and the cargo-free system (extended structure, PDB ID code 4FGV, presented here), a common subset of C_α atoms was selected from residues that are identical among the three orthologues, which was determined by sequence alignment and restricted to α-helical regions of the extended conformation.

ACKNOWLEDGMENTS. We thank Jil Schrader, Maren Müller, and Stephanie Schell for excellent technical assistance; Ralph Kehlenbach for critical reading of the manuscript; and Lin-Ta Hsu for creating the schematic of the transport cycle. This work was supported by Deutsche Forschungsgemeinschaft Sonderforschungsbereich 860 (to R.F.), a scholarship from Boehringer Ingelheim Fonds (to D.H.), the Scottish Universities Physics Alliance (U.Z.), and the UK National Physical Laboratory (U.Z.). Synchrotron data collection was supported by the Helmholtz-Zentrum.

1. Cook A, Bono F, Jinek M, Conti E (2007) Structural biology of nucleocytoplasmic transport. *Annu Rev Biochem* 76:647–671.
2. Görlich D, Kutay U (1999) Transport between the cell nucleus and the cytoplasm. *Annu Rev Cell Dev Biol* 15:607–660.
3. Fornerod M, Ohno M, Yoshida M, Mattaj IW (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* 90(6):1051–1060.
4. Stade K, Ford CS, Guthrie C, Weis K (1997) Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* 90(6):1041–1050.
5. Gadal O, et al. (2001) Nuclear export of 60s ribosomal subunits depends on Xpo1p and requires a nuclear export sequence-containing factor, Nmd3p, that associates with the large subunit protein Rpl10p. *Mol Cell Biol* 21(10):3405–3415.
6. Moy TI, Silver PA (1999) Nuclear export of the small ribosomal subunit requires the ran-GTPase cycle and certain nucleoporins. *Genes Dev* 13(16):2118–2133.
7. Kehlenbach RH, Dickmanns A, Gerace L (1998) Nucleocytoplasmic shuttling factors including Ran and CRM1 mediate nuclear export of NFAT In vitro. *J Cell Biol* 141(4): 863–874.
8. Güttler T, et al. (2010) NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. *Nat Struct Mol Biol* 17(11):1367–1376.
9. Monecke T, et al. (2009) Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* 324(5930):1087–1091.
10. Dong X, et al. (2009) Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature* 458(7242):1136–1141.
11. Paraskeva E, et al. (1999) CRM1-mediated recycling of snurportin 1 to the cytoplasm. *J Cell Biol* 145(2):255–264.
12. Petosa C, et al. (2004) Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol Cell* 16(5): 761–775.
13. Askjaer P, Jensen TH, Nilsson J, Englmeier L, Kjems J (1998) The specificity of the CRM1-Rev nuclear export signal interaction is mediated by RanGTP. *J Biol Chem* 273 (50):33414–33422.
14. Koyama M, Matsuura Y (2010) An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1. *EMBO J* 29(12):2002–2013.
15. Fukuhara N, Fernandez E, Ebert J, Conti E, Svergun D (2004) Conformational variability of nucleo-cytoplasmic transport factors. *J Biol Chem* 279(3):2176–2181.
16. Vetter IR, Arndt A, Kutay U, Görlich D, Wittinghofer A (1999) Structural view of the Ran-Importin beta interaction at 2.3 A resolution. *Cell* 97(5):635–646.
17. Chook YM, Blobel G (1999) Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* 399(6733):230–237.
18. Cook AG, Fukuhara N, Jinek M, Conti E (2009) Structures of the tRNA export factor in the nuclear and cytosolic states. *Nature* 461(7260):60–65.
19. Fox AM, Ciziene D, McLaughlin SH, Stewart M (2011) Electrostatic interactions involving the extreme C terminus of nuclear export factor CRM1 modulate its affinity for cargo. *J Biol Chem* 286(33):29325–29335.
20. Conti E, Müller CW, Stewart M (2006) Karyopherin flexibility in nucleocytoplasmic transport. *Curr Opin Struct Biol* 16(2):237–244.
21. Zachariae U, Grubmüller H (2006) A highly strained nuclear conformation of the exportin Cse1p revealed by molecular dynamics simulations. *Structure* 14(9):1469–1478.
22. Kastner B, et al. (2008) GraFix: Sample preparation for single-particle electron cryomicroscopy. *Nat Methods* 5(1):53–55.
23. Kappel C, Zachariae U, Dölker N, Grubmüller H (2010) An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. *Biophys J* 99(5):1596–1603.
24. Zachariae U, Grubmüller H (2008) Importin-beta: Structural and dynamic determinants of a molecular spring. *Structure* 16(6):906–915.
25. Grinthal A, Adamovic I, Weiner B, Karplus M, Kleckner N (2010) PR65, the HEAT-repeat scaffold of phosphatase PP2A, is an elastic connector that links force and catalysis. *Proc Natl Acad Sci USA* 107(6):2467–2472.
26. Dong X, Biswas A, Chook YM (2009) Structural basis for assembly and disassembly of the CRM1 nuclear export complex. *Nat Struct Mol Biol* 16(5):558–560.
27. Cook AG, Conti E (2010) Nuclear export complexes in the frame. *Curr Opin Struct Biol* 20(2):247–252.
28. Mueller U, et al. (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synchrotron Radiat* 19(pt 3):442–449.
29. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):271–281.
30. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(pt 5):760–763.
31. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(pt 4): 658–674.
32. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(pt 2):213–221.
33. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(pt 2):125–132.
34. Sander B, Golas MM, Stark H (2003) Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J Struct Biol* 142(3):392–401.
35. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* 116(1):17–24.
36. Singer A, Coifman RR, Sigworth FJ, Chester DW, Shkolnisky Y (2010) Detecting consistent common lines in cryo-EM by voting. *J Struct Biol* 169(3):312–322.
37. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
38. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–725.
39. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91:6269–6271.
40. Darden T, York D, Pedersen L (1993) Particle mesh Ewald - an N log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.

Supporting Information

Monecke et al. 10.1073/pnas.1215214110

SI Materials and Methods

Protein Expression and Purification. The cDNA of *Chaetomium thermophilum* chromosome region maintenance 1 (CRM1; *ct*CRM1; exportin1, XPO1) encoding the full-length protein (amino acids 1–1077) was cloned by using EcoRI and NotI restriction sites into a modified pET24d vector. The constructed GST-CRM1 fusion protein contains an interjacent tobacco etch virus (TEV) protease cleavable site and was expressed in *Escherichia coli* BL21(DE3) (Merck). Transformed cells were grown in kanamycin containing 2YT medium supplemented with 2% (wt/vol) α-D-glucose at 37 °C to an OD_{600} of 0.6. Temperature was set to 20 °C and protein expression was induced at an OD_{600} of 0.9 by adding IPTG to a final concentration of 0.6 mM. Cells were harvested after 20 h of induction (5,000 × *g*, 20 min, 4 °C) and resuspended in lysis buffer (500 mM NaCl, 50 mM Hepes/NaOH, pH 7.5, 2 mM $MgCl_2$, and 4 mM DTT). Cells were disrupted using a Microfluidizer 110S (Microfluidics), and the clarified lysate (30,000 × *g*, 30 min, 4 °C) was subsequently loaded onto a GSTrap column (GE Healthcare) equilibrated with lysis buffer. Unbound proteins were removed by washing with two column volumes of lysis buffer, and the bound fusion protein was eluted with a buffer containing additionally 30 mM reduced glutathione. For cleavage, the GST–[TEV]–CRM1 fusion protein was incubated with TEV protease (Invitrogen) at 4 °C overnight in a 1:100 molar ratio of protease:fusion protein. To remove remaining contamination with DNA and/or RNA, 5 μL Benzonase (Merck) was added to the solution. The TEV-cleaved protein was further purified by using a Superdex S200 (26/60) gel filtration column (GE Healthcare) in a buffer containing 150 mM NaCl, 20 mM Hepes/NaOH, pH 7.5, 2 mM $MgCl_2$, and 4 mM DTT. Fractions containing CRM1 were pooled and finally passed over another GSTrap to remove remaining traces of GST and uncleaved fusion protein. Highly pure *ct*CRM1 was concentrated to 20 mg/mL by using Millipore concentrators with a molecular weight cutoff of 30,000 Da (Merck), and 80-μL aliquots were frozen in liquid nitrogen and stored at −80 °C.

Crystallization and Crystal Structure Determination. Purified full-length *ct*CRM1 (amino acids 1–1077) was crystallized in a condition containing 22% (wt/vol) polyacrylic acid sodium salt 5100, 20 mM $MgCl_2$, 100 mM CHES/NaOH, pH 9.0, and 4% (vol/vol) 2,5 hexanediol. Rod-shaped orthorhombic crystals (typically 60 × 60 × 300 μm) grew at 20 °C after 15 d when mixing 1 μL of the reservoir and 1 μL of the protein solution (19 mg/mL). Crystals belong to space group $P2_12_12_1$ with cell dimensions as follows: a, 85.1 Å; b, 139.1 Å; and c, 174.9 Å. The crystals were flash-cooled in liquid nitrogen after soaking in reservoir solution containing, additionally, 14% (vol/vol) glycerol as cryoprotectant.

A diffraction data set has been collected at a wavelength of 0.9184 Å and a temperature of 100 K on BL14.1 operated by the Helmholtz-Zentrum Berlin at the BESSY II electron storage ring (Berlin-Adlershof, Germany) (1). The dataset was integrated and scaled by using iMOSFLM (2) and SCALA (3), respectively.

The structure was solved by means of molecular replacement with PHASER (4) by using the crystal structure of *Mus musculus* CRM1 as search model, which had been extracted from the CRM1–RanGTP–SPN1 complex [Protein Data Bank (PDB) ID code 3GJX] (5). The search was performed with two separate parts of mouse CRM1 (amino acids Ile130-Arg594 and Ala633-Ala1059), which were placed successively by the program. Mouse and *C. thermophilum* CRM1 share a sequence identity and similarity of 50% and 62%, respectively. The sequence was adjusted to *ct*CRM1 and the structure refined by iterative cycles of CNS (6) and manual model building in Coot (7) using a series of mFo-DFc simulated annealing electron density omit maps as implemented in CNS. The final round of refinement was performed with PHENIX (8). The model contains 1,067 of 1,077 residues and was refined at a resolution of 2.94 Å to *R* and R_{free} values of 22.03% and 24.28%, respectively (Table S1). The first two HEAT repeats were poorly defined in the initial 2mFo-DFc map and therefore have been entirely modeled based on simulated annealing omit maps. Model analysis by using MolProbity (9) revealed that there are 94.52% of the residues in the favored, 4.91% in the allowed, and 0.57% in the disallowed regions.

Trigonal crystals of *ct*CRM1 (amino acids 1–1077) grew in a condition containing 22% (wt/vol) polyacrylic acid sodium salt 5100, 20 mM $MgCl_2$, and 100 mM CHES/NaOH, pH 9.0, with a protein concentration of 13 mg/mL. Hexagonal-shaped crystals (typically 100 × 100 × 30 μm) grew at 20 °C after 5 d. Crystals belong to space group $P3_1$ with cell dimensions as follows: a, 89.8 Å; b, 89.8 Å; and c, 316.2 Å. Crystals were flash-cooled according to the protocol for orthorhombic *ct*CRM1, and a diffraction dataset has been collected on beam line P14 at the PETRA III electron storage ring (European Molecular Biology Laboratory, Hamburg, Germany) at a wavelength of 1.2395 Å and a temperature of 100 K. The dataset was integrated and scaled using XDS (10) and XSCALE, and the structure was solved by means of MR (PHASER) using two separate parts of the previously solved orthorhombic *ct*CRM1 structure (amino acids Pro2-Gln641 and Asn643-Lys1068). After careful inspection of the data, merohedral twinning was detected (twin law h,-h-k,-l; twin fraction, 0.46) and implemented in the PHENIX refinement. A gradient map as implemented in CNS for a model missing the acidic loop residues is shown in Fig. S4*B*. There are two CRM1 molecules in the asymmetric unit, and the structure was refined to *R* and R_{free} values of 21.61% and 23.56%, respectively. A total of 96.64% of the residues are located in the favored, 2.87% in the allowed, and 0.49% in the disallowed regions. All structure figures were generated using PyMOL (2002; DeLano Scientific).

EM Preparation and Image Processing. Purified *ct*CRM1 was stabilized using the GraFix protocol (11). The sample was loaded on a 5% to 20% (wt/vol) sucrose gradient in standard buffer containing 0% to 0.1% (vol/vol) glutaraldehyde and was centrifuged in a TH660 rotor (Sorvall) for 20 h at 40,000 rpm at 4 °C. Gradients were fractionated from the bottom, and fixing reaction was quenched by adding 25 mM glycine, pH 7.4. Complexes were then bound to a thin carbon film and transferred to an EM grid covered with a perforated carbon film. The bound molecules were stained with 2% (wt/vol) uranyl formate and air-dried. Images were recorded at a magnification of 155,000× on a 4k × 4k CCD camera (TVIPS) using twofold pixel binning (1.85 Å/pixel) in a CM200 FEG electron microscope (Philips/FEI) operated at 160 kV acceleration voltage (Fig. S8*B*).

A total of 34,000 particle images (144 × 144 pixels) were selected with the custom-written software johnhenry (manuscript in preparation). All valid particles were used to obtain the final result. None of the selected particles was excluded. A total of 13,017 particles were assigned to the compact conformation, and 20,952 particles were assigned to the extended conformation. Contrast transfer function correction was performed on the particle

images based on their classified power spectra (12). Further image processing was done in Imagic (13).

Contrast transfer function-corrected images were coarsened by a factor of two to a pixel size of 3.7 Å/pixel, and reference free-alignment was performed. After several rounds of exhaustive multireference alignment based on resampling in polar coordinates (14) and multivariate statistical analysis (15), stable class averages were obtained (Fig. S7 *A* and *B*). Starting models were generated by using angular reconstitution facilitated by a voting algorithm (16). The handedness was determined from the solved crystal structures. Two distinct conformations could be detected in the dataset, and it was split based in the cross correlation toward both models. To crosscheck those two models, angles for class averages from one model were determined using the other model (Fig. S8*A*). By iterating this procedure, models relaxed to their initial state. Models were further refined to yield a resolution of ~20 Å by projection matching. Resolution was judged by Fourier shell correlation using a cutoff of 0.5 (Fig. S7 *C* and *D*).

Molecular Dynamics Simulations. All simulations were carried out using Gromacs 4.5 (17, 18). The Amber99sb force field (19) was used together with the extended simple point charge water model (20). Electrostatic interactions were calculated using particle-mesh Ewald (21), with a real space cutoff of 1 nm, a grid spacing of 0.13 nm, and cubic interpolation. van der Waals interactions were cut off at a distance of 1.6 nm. Nonbonded interactions were calculated by using neighbor lists, which were updated every five time steps. All simulations were performed in the constant particle number, pressure, and temperature ensemble, using the velocity rescaling method for temperature coupling (22) with a heat bath temperature T of 300 K, a coupling time constant of 0.1 ps, and Berendsen pressure coupling (23), with a reference pressure of 1,000 hPa and a respective coupling time constant of 1 ps. All systems were simulated in dodecahedron-shaped simulation boxes using periodic boundary conditions. All bond lengths were constrained using the LINCS algorithm (24), meaning that harmonic pair bond interactions were removed from the force field potential and replaced by constraints. Additionally degrees of freedom representing fast fluctuations of angles defined by at least one hydrogen atom were removed by changing the respective hydrogen atoms into virtual sites and fixing the angle (25). The equations of motion were integrated using the Verlet algorithm (26) with a time step of 4 fs. Sodium chloride ions were added to ensure neutrality of the simulation box.

System Setup. Four simulation systems based on the extended conformation crystal structure of CRM1 (PDB ID code 4FGV) were constructed: a WT system, a system with a deleted C-terminal helix and the adjacent tail, a system with a deleted acidic loop, and a system containing both deletions. For the C-terminal deletion mutant, all residues starting at Glu1039 were removed. For the acidic loop deletion mutant, residues Glu419 to Asp446 were removed and replaced by a "GGSGGSG" motif, similar to that reported by Lee et al. (27), using Modeler 9v8 software (28).

Five and ten 100-ns unbiased simulations were performed for the WT and double deletion mutant, respectively. For the C-terminal deletion mutant, ten 200-ns simulations were carried out, and, for the acidic loop deletion mutant, five 100-ns simulations were performed.

Before all simulations, an energy minimization of the force field potential was performed by steepest descent, until the step size reached single point precision.

Data Analysis. To compare trajectories from the three different CRM1 structures [the CRM1–RanGTP–SPN1 complex (compact structure, PDB ID code 3GJX), the CRM1–SPN1 complex (almost compact structure, PDB ID code 3GB8), and free CRM1 (extended structure, PDB ID code 4FGV)], a common subset of C_α atoms was selected from residues that are identical in these three proteins. They were determined from a sequence alignment and restricted to the α-helical regions of the open conformation. This set of joint C_α atoms was used to define the linear vectors that connect the three crystal structures and the subsequent construction of the 2D plane in the configurational space that is defined by the three crystal structures. Furthermore, it was used as the subset for a principal component analysis (PCA) of the collected simulations of the mutant with the deleted C-terminal helix, as well as for the projections of the crystal structures and the simulated trajectories onto those vectors.

Nuclear Export Signal Cleft Analysis. For analysis of the behavior of the nuclear export signal (NES) cleft, it was defined to consist of the residues Glu507 to Met527 and Trp556 to Met570 of the extended structure. The C_α atoms of this subset, together with their corresponding counterparts in the compact and almost compact structure, were used to calculate the plane in configurational space defined by the NES cleft atoms of the three crystal structures. For each point in time in every trajectory, the NES cleft was defined to be closed if the Gaussian filtered projection onto the plane defined by the NES configurations in the three crystal structures was closer to the 4FGV structure than to the other structures. Similarly, for each time step, the overall conformation was assigned to an extended or compact ensemble. For this four-state model, the individual probabilities and conditional probabilities of the NES cleft being in the open conformation with the overall structure being extended or compact were calculated. The obtained probability p_i is assumed to be normally distributed with an SD of σ_i being equal to $p_i\,(1 - p_i) / n$, being the effective number of independent samples estimated from the autocorrelation time. A hypothesis test against the null hypothesis, i.e., $P\ (\mathrm{NES_{open}} \mid \mathrm{CRM1_{compact}}) = P\ (\mathrm{NES_{open}} \mid \mathrm{CRM1_{extended}})$ yields significances for these results of 2.3 and 2.6 SDs for the helix deletion and the double deletion mutant, respectively. To crosscheck the obtained probabilities and significances, the transition rates among the four states were estimated from the trajectories. Using these rates, artificial random trajectories were generated in a bootstrap-like approach. Mean and SD of the bootstrap distribution were then used to estimate the conditional probabilities and the corresponding errors. The approach yielded very similar although slightly different values and significances of 1.8 and 3.4 SDs.

1. Mueller U, et al. (2012) Facilities for macromolecular crystallography at the Helmholtz-Zentrum Berlin. *J Synchrotron Radiat* 19(pt 3):442–449.
2. Battye TGG, Kontogiannis L, Johnson O, Powell HR, Leslie AGW (2011) iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67(pt 4):271–281.
3. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(pt 5): 760–763.
4. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(pt 4):658–674.
5. Monecke T, et al. (2009) Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. *Science* 324(5930):1087–1091.
6. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(pt 5): 905–921.
7. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(pt 4):486–501.
8. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(pt 2): 213–221.
9. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66(pt 1):12–21.
10. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(pt 2):125–132.

11. Kastner B, et al. (2008) GraFix: sample preparation for single-particle electron cryomicroscopy. *Nat Methods* 5(1):53–55.
12. Sander B, Golas MM, Stark H (2003) Automatic CTF correction for single particles based upon multivariate statistical analysis of individual power spectra. *J Struct Biol* 142(3):392–401.
13. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M (1996) A new generation of the IMAGIC image processing system. *J Struct Biol* 116(1):17–24.
14. Sander B, Golas MM, Stark H (2003) Corrim-based alignment for improved speed in single-particle image processing. *J Struct Biol* 143(3):219–228.
15. van Heel M (1984) Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy* 13(1-2):165–183.
16. Singer A, Coifman RR, Sigworth FJ, Chester DW, Shkolnisky Y (2010) Detecting consistent common lines in cryo-EM by voting. *J Struct Biol* 169(3):312–322.
17. Van Der Spoel D, Lindahl E, Hess B (2005) GROMACS: Fast, flexible, and free. *J Comput Chem* 26:1701–1718.
18. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
19. Hornak V, et al. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–725.
20. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91:6269–6271.
21. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N Log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
22. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101–014101.
23. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1983) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81:3684–3684.
24. Hess B, Bekker H, Berendsen H, Fraaije J (1997) LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18:1463–1472.
25. Feenstra K, Hess B, Berendsen H (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J Comput Chem* 20: 786–798.
26. Verlet L (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys Rev* 159:98–103.
27. Lee BJ, et al. (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* 126(3):543–558.
28. Sali AA, Blundell TLT (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779–815.

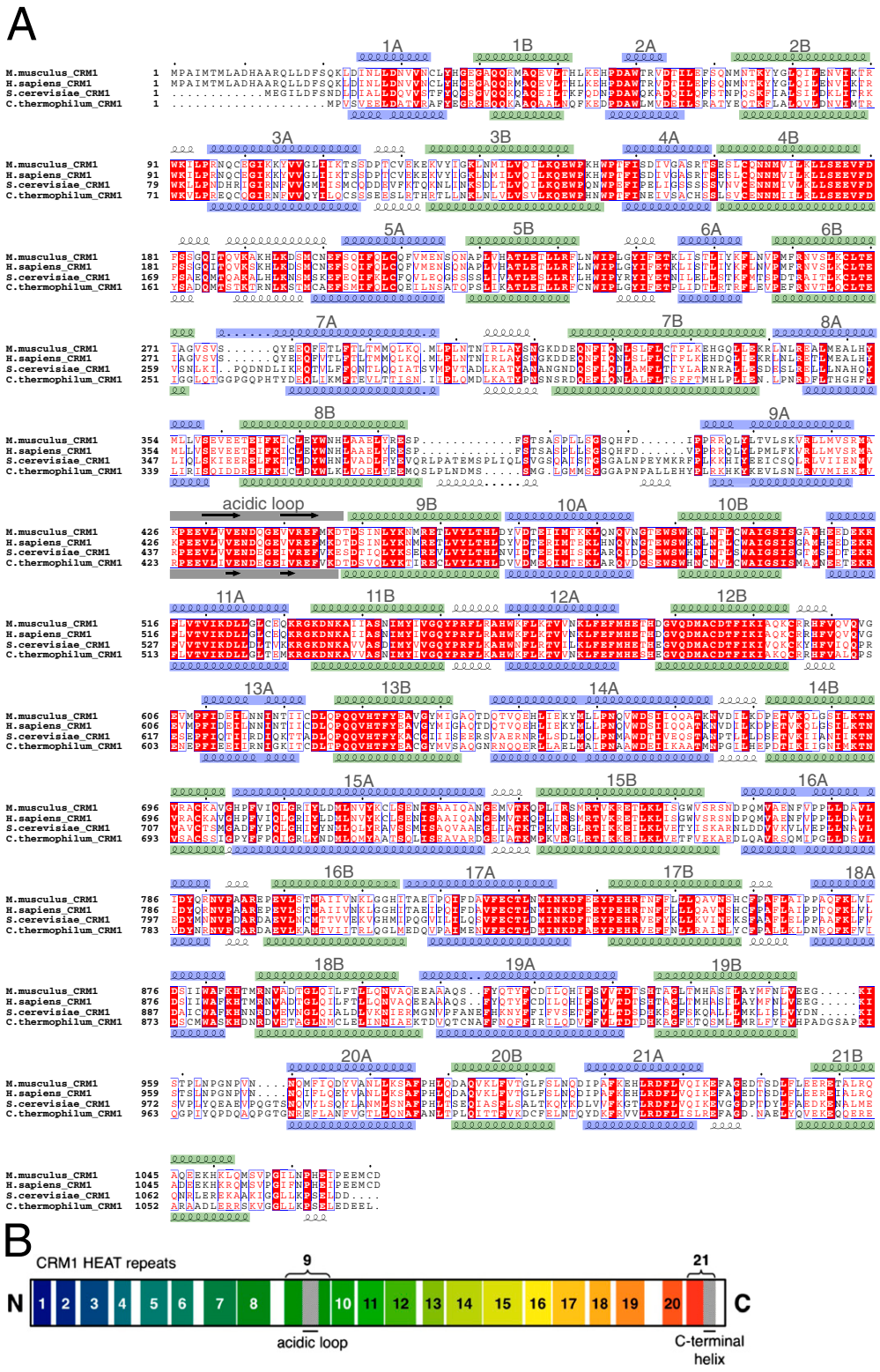


Fig. S1. Sequence alignment and schematic representation of *ct*CRM1. (*A*) Sequence alignment of mouse (*Mus musculus*), human (*Homo sapiens*), yeast (*Saccharomyces cerevisiae*) and *C. thermophilum* CRM1. Secondary structure elements are shown on top of the alignment for mouse CRM1 (PDB ID code 3GJX) and on the bottom for *ct*CRM1 (PDB ID code 4FGV). A-helices (blue) and B-helices (green) of the respective HEAT repeats align quite well and are labeled above the upper secondary structure assignment. *ct*CRM1 encompasses 1,077 aa and shows 62% sequence similarity (50% identity) to the human and 65% similarity (52% identity) to the yeast orthologue. (*B*) Schematic organization of HEAT repeats within *ct*CRM1. The HEAT repeats are numbered and colored from the N terminus (blue) to the C terminus (red). The acidic loop and the C-terminal helix are labeled.

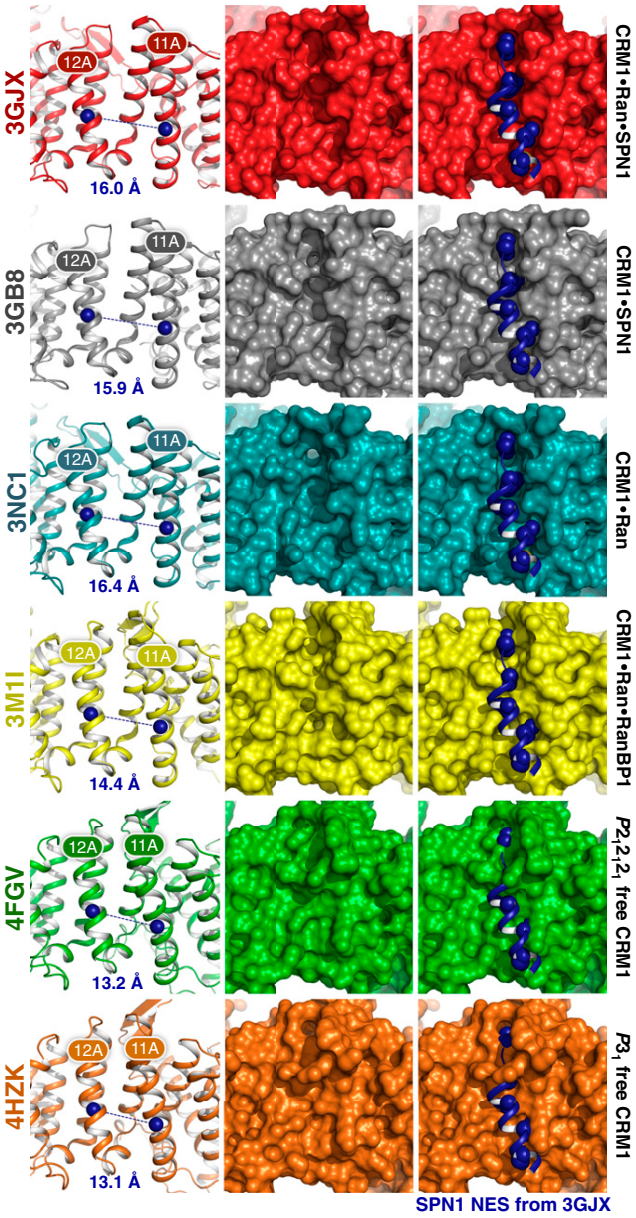


Fig. S2. Detail views of the NES cleft in different CRM1 crystal structures. A front view of the NES cleft is shown as a cartoon image (*Left*) with the centers of mass of the helices 11A and 12A represented by blue spheres. The distances between the centers of mass are denoted below. A surface model (*Middle*) illustrates the differences of the NES clefts between structures of CRM1 bound to SPN1 and RanGTP (PDB ID code 3GJX; red), CRM1 bound to SPN1 (PDB ID code 3GB8; gray), CRM1 bound to RanGTP (PDB ID code 3NC1; deep teal), CRM1 bound to RanGTP and RanBP1 (PDB ID code 3M1I; yellow) and free CRM1 in the orthorhombic (PDB ID code 4FGV; green) and the trigonal form (PDB ID code 4HZK; orange). A superposition of the individual models with the SPN1 NES from the ternary CRM1–RanGTP–SPN1 complex (*Right*) highlights the structural changes in the NES cleft, which are incompatible with NES binding in the free and RanBP1-bound CRM1 structures.

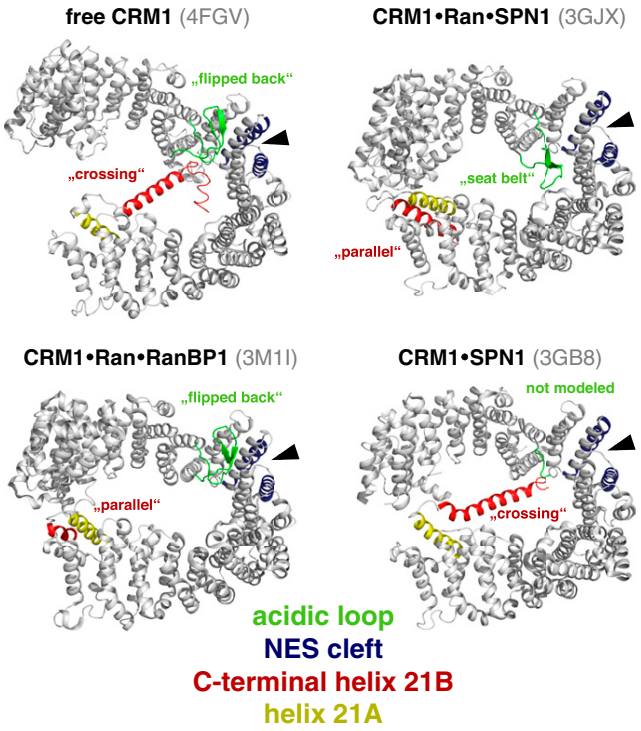


Fig. S3. Conformation and position of the acidic loop (green), the NES cleft (blue), helix 21A (yellow), and the C-terminal helix 21B (red) in different CRM1 crystal structures. For clarity reasons, only CRM1 is depicted in cartoon mode (light gray). Free CRM1 (*Upper Left*; PDB ID code 4FGV), CRM1–RanGTP–SPN1 (*Upper Right*; PDB ID code 3GJX), CRM1–RanGTP–RanBP1 (*Lower Left*; PDB ID code 3M1I), and CRM1–SPN1 (*Lower Right*; PDB ID code 3GB8) are shown. The acidic loop conformations are "flipped back" when bound to the back side of the NES cleft (black arrowhead), like in 4FGV and 3M1I; and in a "seatbelt" conformation when fixed by Ran and contacting the opposite side of CRM1 (like that of PDB ID code 3GJX). The conformation of the C-terminal helix is named "crossing" when it diagonally crosses the CRM1 arch and contacts a basic patch on the opposite side of CRM1 (like in PDB ID codes 3GB8 and 4FGV) and "parallel" when helix 21B is aligned in parallel to helix 21A (like in PDB ID codes 3GJX and 3M1I).

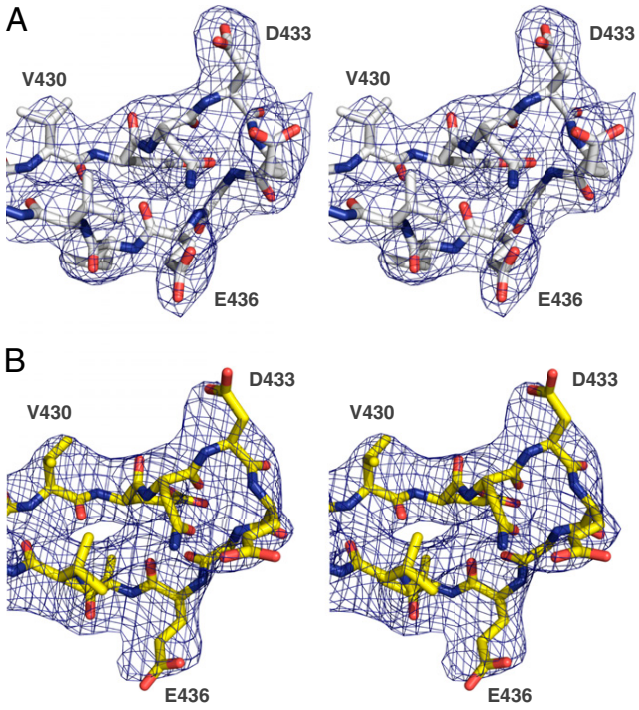


Fig. S4. Conformations and electron density omit maps of the acidic loop in free *ct*CRM1 crystal structures. (*A*) Stereo image of a representative mFo-DFc simulated annealing electron density omit map (residues 428–440 omitted) of the acidic loop from the orthorhombic crystal form (PDB ID code 4FGV). (*B*) For the trigonal crystal form, a stereo image of a representative electron density gradient map for a model lacking the acidic loop residues 428 to 440 was calculated as implemented in PHENIX (PDB ID code 4HZK). Stick models of the acidic loop residues are shown, and both electron density maps are contoured at a σ-level of 3.0.

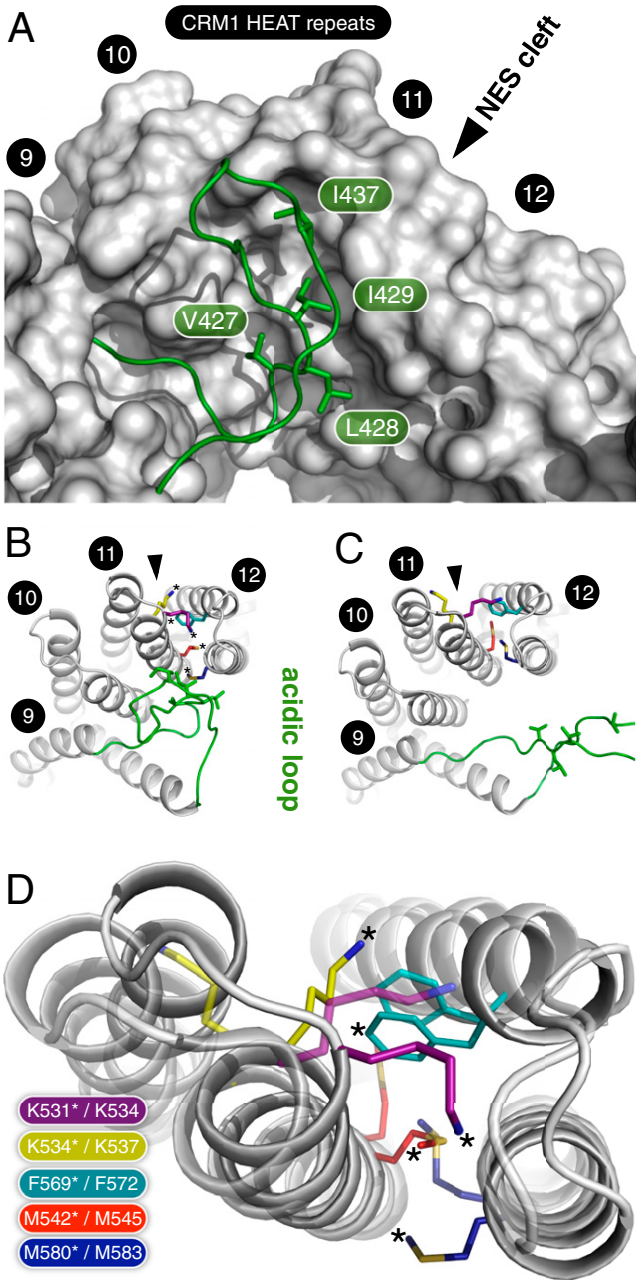


Fig. S5. Interactions of the acidic loop with the back side of the NES cleft stabilize its closed conformation. (*A*) Stabilization of the NES cleft in the closed conformation is achieved by hydrophobic interactions between residues of the acidic loop (green) and the area formed by the B-helices of HEAT repeats 11 and 12. HEAT repeats and residues of the acidic loop are labeled. Below the rearrangement of residues within the hydrophobic core of the helix bundle forming the NES cleft is shown for the extended conformation (PDB ID code 4FGV) (*B*) and the compact conformation (PDB ID code 3GJX) (*C*). The corresponding residues of the two structures are numbered and labeled in identical colors. (*D*) Superposition of only the prominent residues from *B* and *C* with the respective residues in identical colors. The rearrangements from the extended form (asterisk; PDB ID code 4FGV) to the compact conformation (transparent; PDB ID code 3GJX) are shown.

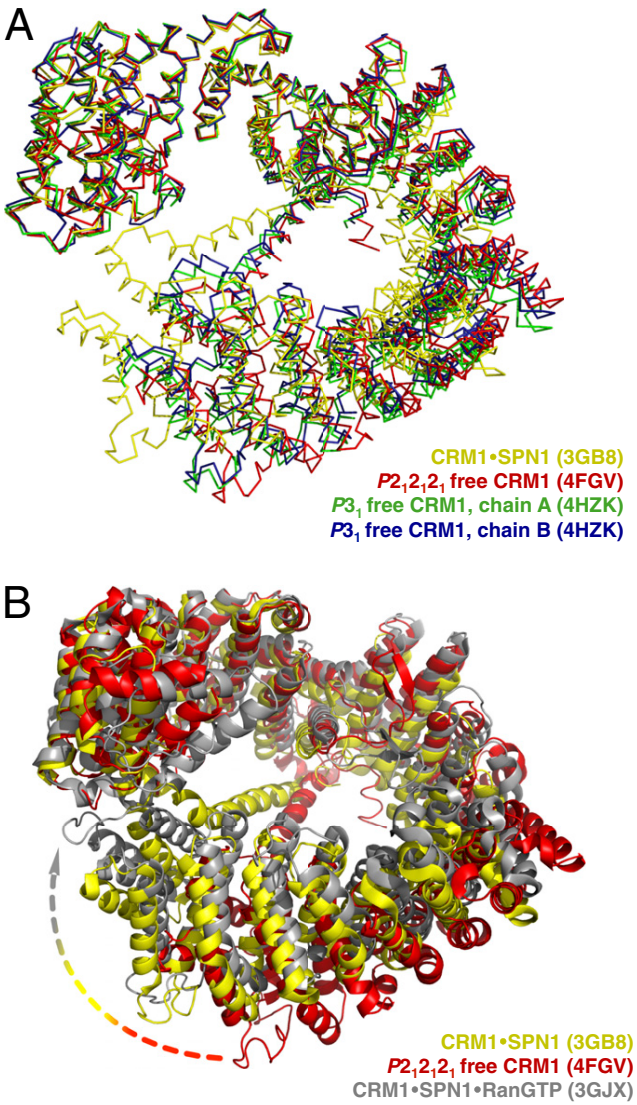


Fig. S6. Superposition of the two structures of free *ct*CRM1 ($P2_12_12_1$ and $P3_1$) and CRM1 bound to SPN1 (yellow) (*A*). For clarity reasons, only CRM1 is shown as C_α trace. Orthorhombic free CRM1 is shown in red, and the two molecules of the trigonal crystal form of free CRM1 are depicted in green and blue, respectively. Note the high degree of superposition of the free CRM1 structures in the N-terminal half of the molecule, restricting main structural changes to the C-terminal arch. (*B*) Superposition of free *ct*CRM1 (extended conformation, red), CRM1 bound to SPN1 (almost compact conformation, yellow), and CRM1 bound to RanGTP and SPN1 (compact conformation, gray).

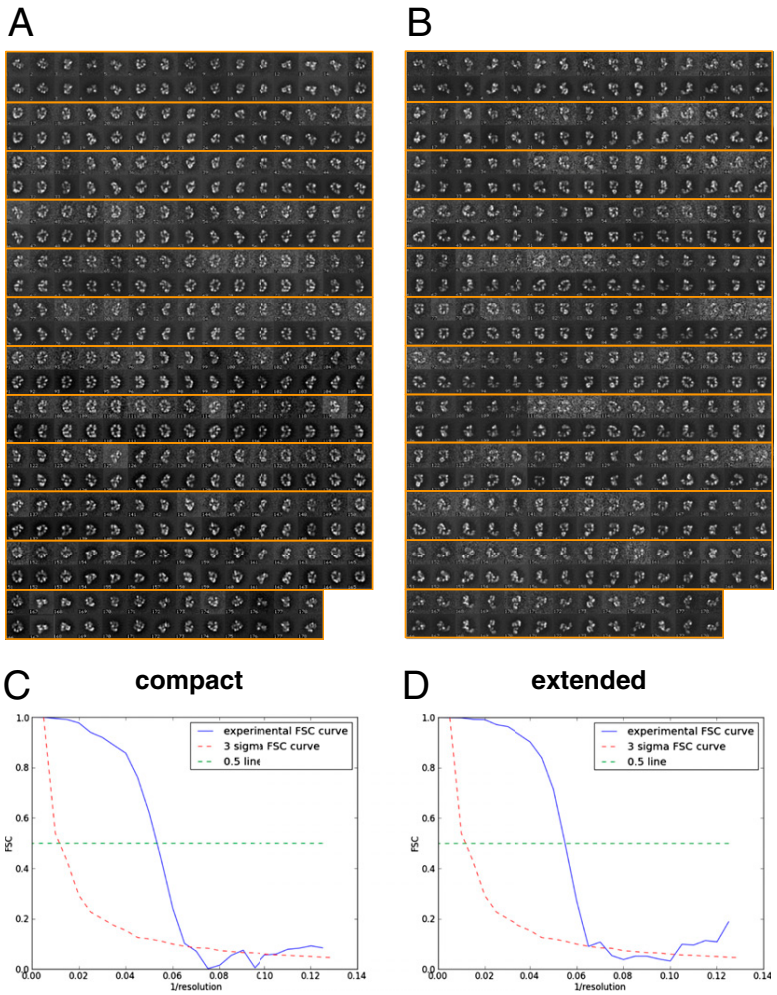


Fig. S7. EM data analysis. Reconstruction of the compact (*A*) and extended (*B*) conformation of *ct*CRM1. The particle classes 1 to 178 for each reconstruction are shown. Each orange framed row contains the particle sum (upper line) as well as the respective back projection (lower line) of a given particle class. Fourier shell correlation (FSC) curves for the compact (*C*) and extended (*D*) EM density maps were used to approximate the resolution of the respective models.

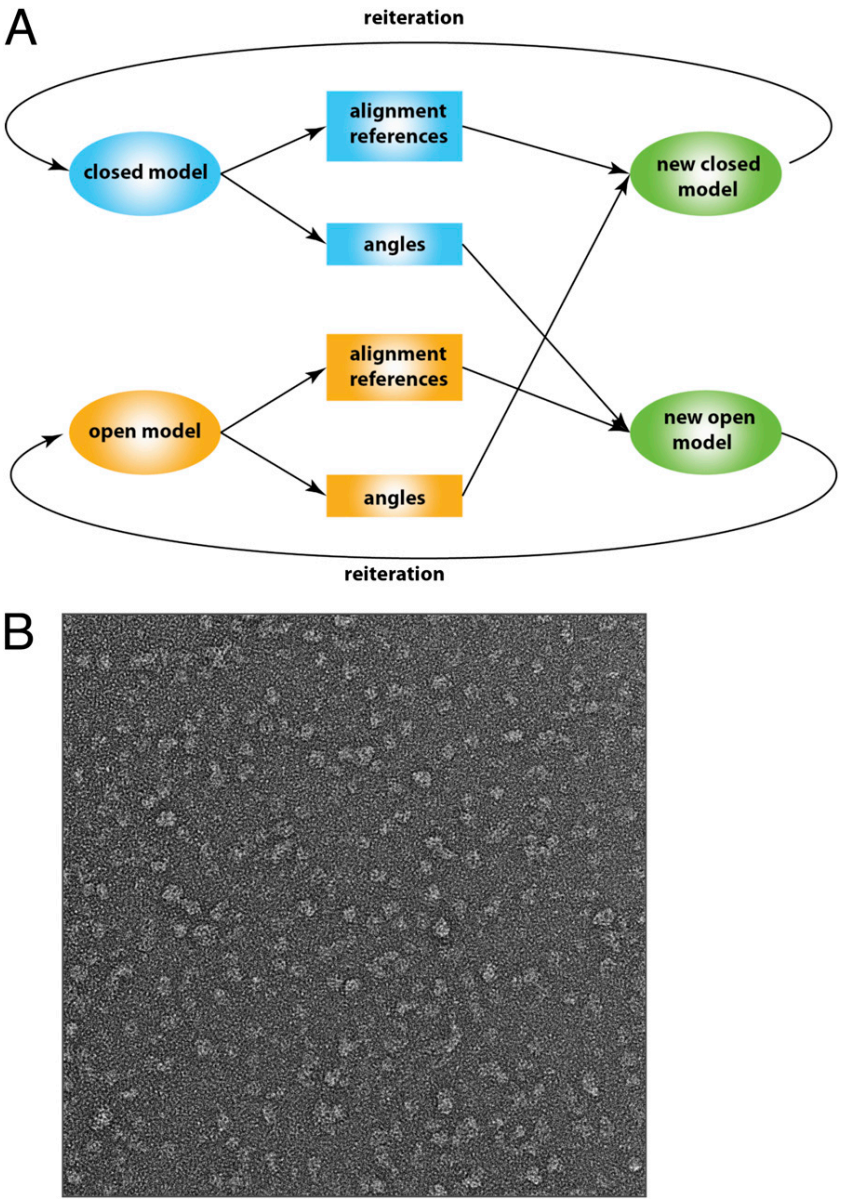


Fig. S8. Scheme for the EM data validation procedure used (*A*) and a representative EM raw image (*B*).

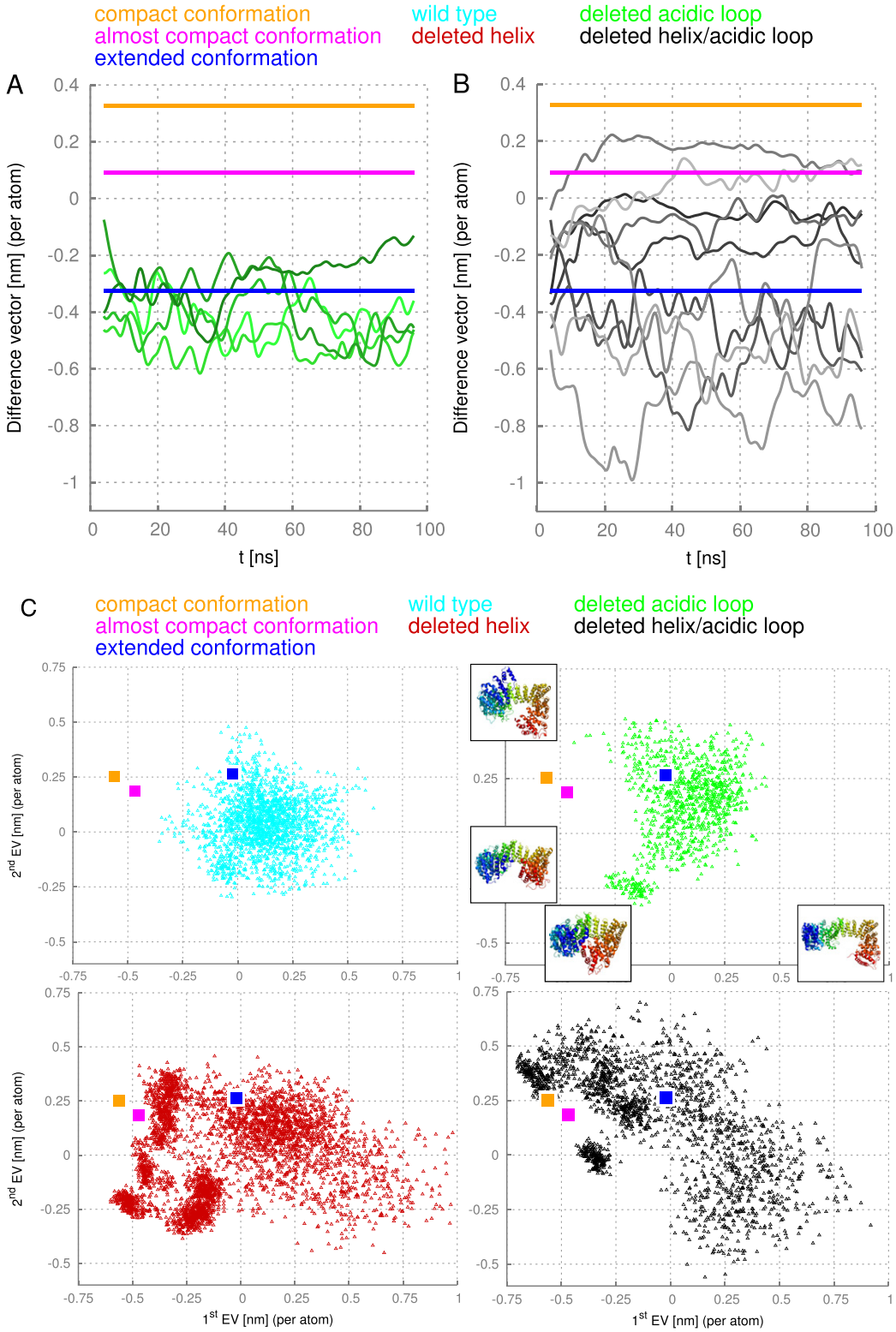


Fig. S9. Projections of the acidic loop deletion mutant (*A*, green) and the double deletion mutant (*B*, acidic loop and C-terminal helix deleted, gray) onto the difference vector between the extended (blue) and compact (orange) CRM1 structure. Note that deletion of the acidic loop alone does not change the structure significantly in molecular dynamics simulations. Additional deletion of the C-terminal helix increases the structural flexibility of the molecule and leads to a more compact conformation of CRM1. (*C*) Projections of WT and all mutant simulations onto the first two eigenvectors of a PCA of the helix deletion mutant simulation. The projections show dynamic movements during simulations along the second eigenvector that are not contained by the three crystal structures.

Table S1. X-ray data collection, refinement, and validation statistics for both crystal forms of full-length *C. thermophilum* CRM1

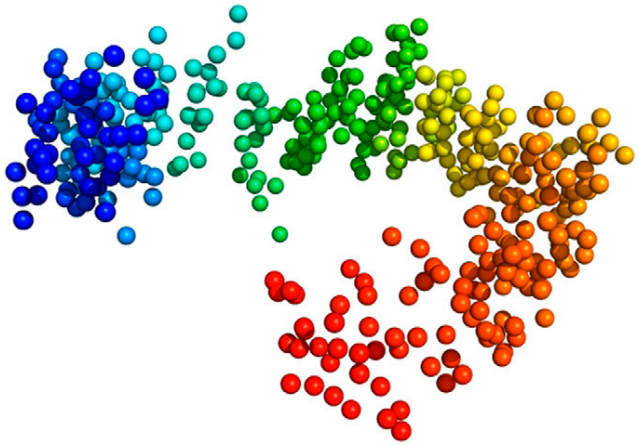
Variable	CRM1; BL14.1, BESSY II	CRM1; P14, PETRA III
Data collection		
Space group	$P2_12_12_1$	$P3_1$
Number of molecules/a.u.	1	2
Wavelength, Å	0.9184	1.2395
Cell dimensions		
a, *b*, *c*; Å	85.12, 139.07, 174.87	89.82, 89.82, 316.20
α, β, γ; °	90.00, 90.00, 90.00	90.00, 90.00, 120.00
Resolution, Å	51.44–2.94 (3.10–2.94)	50.00–3.10 (3.20–3.10)
R_{merge}	0.087 (0.545)	0.042 (0.562)
$I/\sigma(I)$	9.6 (2.1)	17.23 (2.18)
Completeness, %	96.4 (84.5)	96.9 (98.5)
Multiplicity	3.6 (3.3)	3.2 (3.2)
Refinement		
Resolution, Å	45.85–2.94	49.07–3.10
No. reflections	43,007	50,204
R_{work}	0.220	0.216
R_{free}	0.243	0.236
No. atoms	8,595	16,642
B-factors, Å^2	93.06	110.93/113.28 (chain A/B)
rmsd		
Bond lengths, Å	0.004	0.004
Bond angles, °	0.884	0.867
PDB ID code	4FGV	4HZK

All values in parentheses are for the respective highest resolution shell.

Table S2. Conditional probabilities of observing the NES cleft in the open conformation

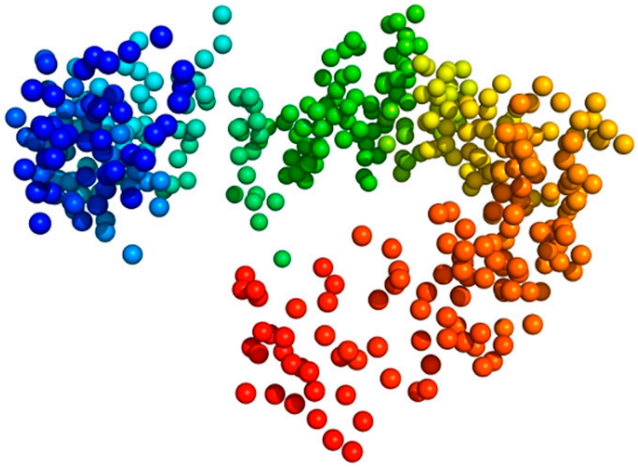
Variable	Deleted helix 21B	Deleted acidic loop + helix 21B
P value		
NES_{open} vs. $CRM1_{extended}$	0.044 ± 0.021	0.070 ± 0.036
NES_{open} vs. $CRM1_{compact}$	0.121 ± 0.033	0.235 ± 0.060
Increase	0.023	0.010
Corresponding ΔΔG, kJ/mol	2.7 ± 1.4	3.5 ± 1.6

The probability of finding the NES cleft in an open conformation is increased by a factor of approximately three if CRM1 adopts the almost compact conformation compared with CRM1 in the extended conformation. This constitutes a significant shift in the free energy difference between the open and closed NES cleft conformation.



Movie S1. Main dynamics along eigenvector 1. Movements of *ct*CRM1 during the simulations along the first eigenvector of the PCA (Fig. S9*C*) constitutes solely a change in the diameter of the ring. The spheres represent a common subset of C_α atoms selected from residues that are identical in human, mouse, and *C. thermophilum* CRM1. Methodological details are provided in *SI Materials and Methods*.

Movie S1



Movie S2. Main dynamics along eigenvector 2. Movements of *ct*CRM1 during the simulations along the second eigenvector of the PCA (Fig. S9*C*) constitutes solely a change in the pitch of the superhelix. The spheres represent a common subset of C_α atoms selected from residues that are identical in human, mouse, and *C. thermophilum* CRM1. Methodological details are provided in *SI Materials and Methods*.

Movie S2

Cell PRESS

Structure
Article

Structural Determinants and Mechanism of Mammalian CRM1 Allostery

Nicole Dölker,[1,8,9] Clement E. Blanchet,[3,8] Béla Voß,[1,8] David Haselbach,[2,8] Christian Kappel,[1,10] Thomas Monecke,[4] Dmitri I. Svergun,[3] Holger Stark,[2,5] Ralf Ficner,[4] Ulrich Zachariae,[6,7] Helmut Grubmüller,[1,*] and Achim Dickmanns[4,*]
[1]Abteilung für Theoretische und Computergestützte Biophysik
[2]Dreidimensionale Kryo-Elektronenmikroskopie
Max-Planck-Institut für Biophysikalische Chemie, Am Faßberg 11, 37077 Göttingen, Germany
[3]European Molecular Biology Laboratory, Hamburg Unit, EMBL c/o DESY, Notkestraße 85, 22603 Hamburg, Germany
[4]Abteilung für Molekulare Strukturbiologie
[5]Abteilung für Molekulare Kryo-Elektronenmikroskopie
Institut für Mikrobiologie und Genetik, Göttinger Zentrum für Molekulare Biowissenschaften (GZMB), Georg-August-Universität Göttingen, Justus-von-Liebig-Weg 11, 37077 Göttingen, Germany
[6]Division of Computational Biology, College of Life Sciences
[7]Division of Physics, School of Engineering, Physics and Mathematics
University of Dundee, Dundee DD1 5EH, UK
[8]These authors contributed equally to this work
[9]Present address: Centro Nacional de Investigaciones Oncológicas, C/ Melchor Fernández Almagro, 3, 28029 Madrid, Spain
[10]Present address: Konrad-Adenauer-Straße 75, 69207 Sandhausen, Germany
*Correspondence: hgrubmu@gwdg.de (H.G.), adickma@gwdg.de (A.D.)
http://dx.doi.org/10.1016/j.str.2013.05.015

SUMMARY

Proteins carrying nuclear export signals cooperatively assemble with the export factor CRM1 and the effector protein RanGTP. In lower eukaryotes, this cooperativity is coupled to CRM1 conformational changes; however, it is unknown if mammalian CRM1 maintains its compact conformation or shows similar structural flexibility. Here, combinations of small-angle X-ray solution scattering and electron microscopy experiments with molecular dynamics simulations reveal pronounced conformational flexibility in mammalian CRM1 and demonstrate that RanGTP binding induces association of its N- and C-terminal regions to form a toroid structure. The CRM1 toroid is stabilized mainly by local interactions between the terminal regions, rather than by global strain. The CRM1 acidic loop is key in transmitting the effect of this RanGTP-induced global conformational change to the NES-binding cleft by shifting its population to the open state, which displays enhanced cargo affinity. Cooperative CRM1 export complex assembly thus constitutes a highly dynamic process, encompassing an intricate interplay of global and local structural changes.

INTRODUCTION

In contrast to prokaryotic cells, eukaryotic cells reveal a high degree of spatial compartmentalization into membrane-engulfed entities. This, for instance, enables a strict spatiotemporal separation of cellular processes such as transcription, occurring in the nucleus, and translation in the cytoplasm. Transport between the nucleus and the cytoplasm proceeds through nuclear pore complexes (NPC) and depends on specialized transport systems. Macromolecules exceeding 30–40 kDa require the aid of karyopherins (KAPs) as mediators to pass the NPC efficiently (Chook and Süel, 2011; Cook and Conti, 2010).

The majority of KAPs are members of a superfamily named after Importin-β (Impβ), the first receptor identified (Görlich et al., 1997; Radu et al., 1995). They are divided into importins and exportins according to the direction of cargo transport. Their common biochemical properties are the capability to interact with the NPC and bind to the small GTPase Ran (Ras-related nuclear antigen). The asymmetric distribution of the Ran-regulating factors with the Ran guanine-nucleotide exchange factor (RanGEF) residing in the nucleus and the Ran GTPase activating protein (RanGAP) located in the cytoplasmic compartment ensures that nuclear Ran predominantly occurs in its GTP-bound form. In contrast to the cytoplasmic, GDP-bound form of Ran, RanGTP can bind to KAPs. RanGTP binding modulates the affinity of KAPs for cargo and thereby enforces directionality of transport.

On a structural level, all members of the Impβ superfamily share a common arrangement of about 20 building blocks, so-called HEAT repeats (Kobe et al., 1999), each consisting of two antiparallel α helices connected by a loop. Their consecutive arrangement results in an overall superhelical shape resembling a solenoid (Fontes et al., 2000). In exportins, RanGTP promotes cargo binding predominantly by interacting simultaneously with receptor and cargo, as for instance seen in Exportin-t, Exportin5, or Cse1p/CAS (Cook et al., 2005, 2009; Matsuura and Stewart, 2004; Okada et al., 2009). In contrast, the export receptor CRM1 (chromosome region maintenance 1), which recognizes the majority of proteins destined for export (Hutten and Kehlenbach, 2007), displays no direct interaction of Ran and cargo. CRM1 in the cargo-bound state exhibits a toroidal, compact,

 ©2013 Elsevier Ltd All rights reserved

Structure
Structural Insight into Mammalian CRM1 Cooperativity

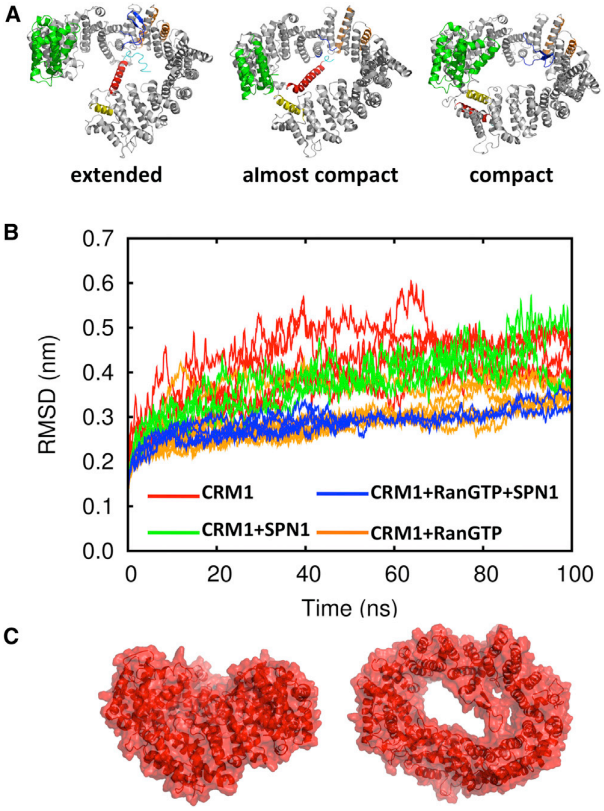


Figure 1. Changes of the Overall Structure of CRM1 during MD Simulations

(A) Crystal structures showing the three most prominent conformations of CRM1. Left: Extended conformation as in free CRM1 (4FGV) with no interaction between N- (green) and C-terminal regions (HEAT helix 21A, yellow); the AL (blue) in the flipped back position and the HEAT helix 21B (red) in the bridging position. The NES binding cleft is shown in orange. In the almost compact conformation as in the CRM1-SPN1 complex (3GB8), few interactions between N- and C-terminal regions are seen, the AL is not resolved, and helix 21B is in the bridging position, but exhibits a kink. In the compact conformation as in the CRM1-RanGTP-SPN1 complex (3GJX), close contacts between N- and C-terminal regions are seen, the AL is in the seatbelt conformation, and helix 21B is in the parallel position on the outside of the CRM1 molecule.

(B) Structural changes of CRM1 in the ternary complex during MD equilibration, monitored by the RMSD to the crystal structure (3GJX; blue curves). Changes in the rmsd of CRM1 in complex with either RanGTP or SPN1 are shown in orange or green, respectively; the red curves represent changes for CRM1 alone.

(C) CRM1 maintains a toroidal structure during MD simulations as shown by a snapshot of CRM1 in the free form after a 100-ns simulation (bottom right).

See also Figures S1 and S2 and Tables S1 and S2 for additional information.

shape with the N- and C-terminal HEAT repeats in close contact (Koyama and Matsuura, 2010; Monecke et al., 2009). A coexisting less compact but still toroidal shape has been described during some states of its transport cycle (Dong et al., 2009b; Figure 1A). Recent structural analysis of free CRM1 from *Chaetomium thermophilum* (ctCRM1) and *Saccharomyces cerevisiae* (scCRM1) revealed that, in these organisms, CRM1 also adopts a more or less extended superhelical shape without close interaction of the N- and C-terminal regions (Monecke et al., 2013; Saito and Matsuura, 2013).

Crystal structures of various CRM1 complexes have provided insight into molecular details of the interactions between CRM1 and its interaction partners during the transport cycle. CRM1 cooperatively binds RanGTP and cargo in the nucleus (Paraskeva et al., 1999). In this ternary complex, RanGTP is localized within the ring of CRM1 and bound by N- and C-terminal HEAT repeats as well as the acidic loop (AL). The AL is inserted between the helices of HEAT repeat 9 (H9) and affixes the GTPase to the terminal HEAT repeats like a seatbelt (Monecke et al., 2009). Remarkably, the cargo binds on the opposite side of CRM1 without direct contacts to RanGTP. It predominantly interacts with acidic patches on the outer surface of CRM1 and a groove formed between the α helices of H11 and H12 (Dong et al., 2009a, 2009b; Monecke et al., 2009). A common motif required for binding of cargo within this groove is the leucine-rich nuclear export signal (NES) consisting of a short peptide stretch of 10–15 residues (Güttler et al., 2010). The CRM1-RanGTP-cargo complex traverses the NPC and enters the cytosol, where it dissociates upon binding of disassembly factors such as RanBP1, which function as cargo release factors and increase the hydrolysis rate of Ran-bound GTP when binding to RanGAP (Askjaer et al., 1999; Koyama and Matsuura, 2010; Maurer et al., 2001; Paraskeva et al., 1999). Free CRM1 shuttles back to the nucleus for another round of export. The reported crystal structures reveal snapshots of various states during the transport process and show the interaction surfaces of CRM1 with cargo and/or Ran. Due to the growing medical interest in CRM1 and its role in cancer (Turner et al., 2012), it is important to understand the dynamics of human CRM1 with a focus on the NES-binding cleft where many therapeutics bind. Purely static structure characterization alone is insufficient for a complete description of structural changes during the transport cycles. Recent findings from MD simulations on the free—extended—form of CRM1 from the lower eukaryote *C. thermophilum,* have shown that the α helix of H21, but not the AL, contribute significantly to the ratio between the extended and the compact form of CRM1 (Monecke et al., 2013). In the ternary complex of CRM1-RanGTP-SPN1, the altered arrangement of the AL, bridging the central opening and linking two distant regions of CRM1, suggests a structural role for determining both the overall conformation of CRM1 and that of the NES-binding cleft. Moreover, the role of RanGTP in restricting the conformational flexibility of CRM1, especially regarding the NES-binding cleft, and the opening mechanism of the toroidal form of CRM1 toward the extended conformation are still open questions. Here, small-angle X-ray scattering (SAXS), electron microscopy (EM), and molecular dynamics (MD) simulations were combined with the available information from crystal structures to elucidate the structural transitions and forces required for the cooperative binding and release of RanGTP and/or the cargo Snurportin1 (SPN1) to mammalian CRM1. We find that mammalian CRM1 in the free form reveals a high degree of conformational flexibility. Binding of RanGTP decreases this flexibility and shifts the conformation toward a more rigid, compact form of CRM1. Our results also show that the AL has a strong influence over the state of the NES-binding cleft. We conclude that RanGTP binding in the presence of the AL ensures

Table 1. Characteristics Determined by SAXS Measurements

	R_g (nm)	D_{max} (nm)	Porod (nm^3)	Estimated Molecular Mass (Porod)	Expected Molecular Weight
mmCRM1	3.8 ± 0.1	11 ± 1	190 ± 20	120 ± 10	121
hsCRM1	3.9 ± 0.1	11 ± 1	180 ± 20	110 ± 10	121
hsCRM1 + SPN1	4.1 ± 0.4	14 ± 1	260 ± 20	160 ± 15	162
mmCRM1 + RanGTP (+NES)	3.6 ± 0.1	10 ± 1	230 ± 20	140 ± 15	141
mmCRM1 + RanGTP + SPN1	4.1 ± 0.1	14 ± 1	300 ± 30	190 ± 20	183

All data were calculated using the programs indicated in the Supplemental Experimental Procedures.
See also Tables S1 and S2.

that the NES-binding cleft for export cargo remains in an open conformation prone for NES binding, and thus enhances the affinity for cargo.

RESULTS AND DISCUSSION

Free MD Simulations of Mammalian CRM1

To gain insight into the atomic rearrangements in mammalian CRM1 during disassembly, we performed multiple unrestrained 100-ns MD simulations of the mouse (mm)CRM1-RanGTP-SPN1 ternary complex (Protein Data Bank [PDB] ID: 3GJX) and on the same assembly structure after removing either SPN1 or/and RanGTP. Global conformational changes were monitored by calculation of the Cα root-mean-square deviation (rmsd) values relative to the crystal structure (Figure 1B). The ternary complex in solution shows a significant increase in the backbone rmsd of CRM1 only in the first 2–5 ns (Figure 1B). This probably reflects a fast adaptation or relaxation from a polyethylene glycol (PEG)-containing condition, in which the crystals were grown, to a PEG-free solution in the MD simulation. Moreover, we considered individual complexes in the simulations, relieving possible strain from crystal contacts. After this initial phase, only a moderate increase of the rmsd is seen during the rest of the simulation. When SPN1 was removed, CRM1 underwent only little additional overall change (Figure 1B), as indicated by the small rmsd increase. The conformational stability of the CRM1-RanGTP-SPN1 and CRM1-RanGTP complexes is also reflected by the radius of gyration of CRM1 (R_g), which remains stable during the simulations (Figure S1 available online).

The overall rmsd of free CRM1 is increased over the ternary complex and stronger fluctuations in R_g are observed (Figures 1B and S1A). The CRM1-SPN1 complex exhibited an intermediate rmsd behavior, increasing more markedly than that of the CRM1-RanGTP-SPN1 complex and reaching the values of free CRM1 at the end (Figure 1B). Overall, the shape of CRM1 stayed ring-like in all simulations (Figure 1C), and the AL remained near the seatbelt conformation observed in the crystal structure (Figures S1B and S2). In all cases, after 100 ns of simulation, the overall rmsd had still not fully converged, indicating that the simulations had not yet reached equilibrium and that further structural rearrangements may take place on a larger timescale.

In contrast to the simulations, the structures of free CRM1 from *C. thermophilum* (4FGV and 4HZK; Monecke et al., 2013) and *S. cerevisiae* (3VYC; Saito and Matsuura, 2013) show CRM1 to adopt a more or less extended superhelical shape, respectively. Because these conformations are not observed in the MD simulations, the question arises whether the extended conformations are specific to CRM1 from lower eukaryotes or if such a conformational change is inaccessible on the time scale of MD simulations. To clarify this question, we performed EM and SAXS experiments to elucidate the global shape and the extent of rearrangement of the AL in mammalian CRM1.

SAXS Measurement, Ab Initio Modeling and Subtractive Modeling of Mammalian CRM1 and Complexes

Human (hs)RanGTP, hsSPN1, hsCRM1, and mmCRM1 (differing only in a few residues; Figure S3) were purified, and the individual complexes were assembled and then analyzed. CRM1 in complex with only RanGTP could not be analyzed due to instability of the complex (Dong et al., 2009b). Thus the complex of CRM1 and RanGTP was stabilized by a short peptide resembling a leucine-rich NES. The SAXS of CRM1, the ternary complexes of CRM1-RanGTP-SPN1 or CRM1-RanGTP-NES as well as CRM1 in complex with SPN1 were measured, and the data were processed, merged, and analyzed (Figure S4A). The Porod volumes and corresponding molecular masses for all samples are consistent with monomeric assemblies in solution (Table 1). The maximum sizes D_{max} and the R_g values (Table 1) were calculated from the distance distribution functions. With use of the range of scattering vectors up to s = 0.2 $Å^{-1}$, low resolution ab initio models of CRM1 alone and of the complexes were constructed (Figures 2 and S4). The ab initio reconstruction from the free CRM1 data yielded a toroidal structure (discrepancy χ = 1.7; Figure 2A). Theoretical scattering patterns of free CRM1 in the extended form (4FGV) and in the compact form, extracted from the ternary complex (3GJX), were computed (see Experimental Procedures). The calculated curves differ significantly from the experimental SAXS results (Figure S5A) so that neither the extended nor the compact form (χ = 3.2 and χ = 3.4, respectively) fit well. Better fits were obtained using free ctCRM1 (4HZK) and scCRM1 (3VYC), which are less extended than CRM1 in PDB ID 4FGV, and with hsCRM1 extracted from PDB ID 3GB8. The better fit for the latter is in agreement with recent results (Fox et al., 2011), but one should note that in all these structures, up to 12% of atoms present in full CRM1 are not resolved (see legend of Table S2).

The CRM1-SPN1 complex reveals a toroidal shape of CRM1 with SPN1 attached on the outside (Figure 2C). The theoretical scattering curve computed from the binary complex extracted from 3GJX shows a significant misfit to the experimental data (χ = 2.8). The conformation in solution revealed by SAXS appears therefore noticeably more extended than the structure observed

Structure
Structural Insight into Mammalian CRM1 Cooperativity

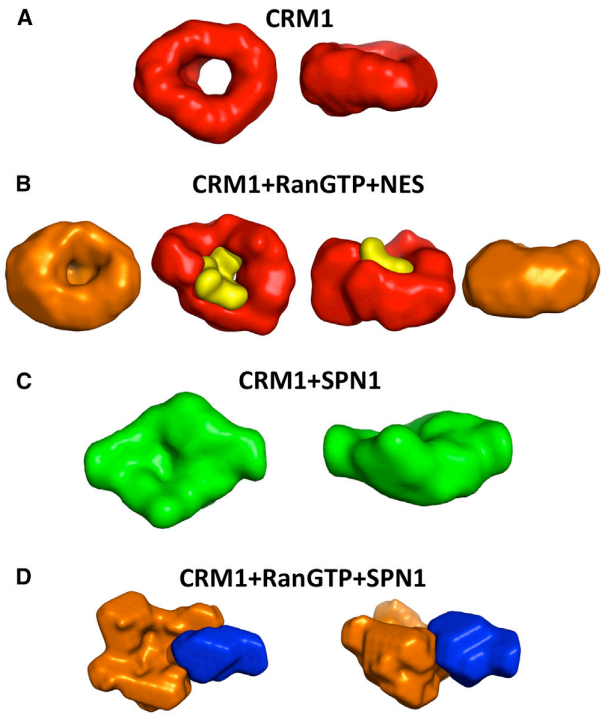


Figure 2. Localization of the Individual Components of the CRM1 Complexes by Comparative Structure Determination Using the Set of SAXS Data Curves

Processed solution scattering patterns from mmCRM1, mmCRM1-hsRanGTP, hsCRM1-hsSPN1, and mmCRM1-hsRanGTP-hsSPN1 (Figures S3–S5) were used to calculate the ab initio models.

(A and B) CRM1 depicted in red (top) reveals a toroidal shape in solution (A) and maintains this shape upon RanGTP binding (B, orange model). Modeling of the individual molecules localizes RanGTP (yellow) in the hollow of CRM1 (red).

(C) SPN1 bound to CRM1 (green model).

(D) SPN1 (blue) clearly localizes to the outer surface of CRM1 (CRM1 and RanGTP in orange).

See also Tables S1 and S2.

in the ternary complex and, given that the domains of SPN1 itself are expected to be rather rigid (Table S1), this points to an extended structure of CRM1 itself. The extended conformation more likely resembles the one observed in the CRM1-SPN1 structure (3GB8).

For CRM1 in complex with RanGTP and NES (Figure 2B) or, RanGTP and SPN1 (Figure 2D), the ab initio models show Ran positioned in the central opening seen for free CRM1. Interestingly, both R_g and D_{max} of CRM1 alone are larger than those for CRM1 in complex with RanGTP, again indicating that CRM1 changes its structure and adopts a more compact conformation upon binding RanGTP. The average R_g values obtained by SAXS are in good agreement with the R_g values for snapshots from the individual MD simulations and the X-ray structures (Table S2).

The overall shape of CRM1 is still recognizable in the ab initio reconstruction derived from the curve of the CRM1-RanGTP-NES complex, but the complex seems to adopt a more compact form. An additional part is observed located close to one side of the ring and to the central opening (discrepancy χ = 2.0; Figure S4C). By simultaneously fitting the experimental curves of the different samples, multiphase ab initio models were built (see Supplemental Experimental Procedures) to gather information of the relative orientation and position of the individual proteins within the complexes. The models fit the experimental data quite well, with a discrepancy of χ = 1.2 for the curve of CRM1 alone and χ = 1.2 for CRM1-RanGTP-NES. The result for the CRM1-RanGTP-NES complex clearly shows CRM1 as a torus with RanGTP in the central opening (Figure 2B). This result is in good agreement with the crystal structure of the CRM1-RanGTP complex (3NC1).

The ab initio structure reconstructed from the SAXS pattern of CRM1-RanGTP-SPN1 clearly adds volume to the outer surface of the CRM1 ring thus localizing SPN1 exactly at this position (discrepancy χ = 1.8; Figures 2D). Due to the fact that the curves obtained for CRM1 alone and in complex with RanGTP differ in R_g and D_{max}, the position of SPN1 in the ternary complex can be determined only with regard to the CRM1-RanGTP complex. In the resulting model (fitting the data with χ = 1.3), SPN1 appears as an appendix attached to the outside of the CRM1-RanGTP-NES shape (Figure 2C).

Taken together, the SAXS data strongly indicate that unbound mammalian CRM1 exists in a more extended structure in the measured ensemble of molecules and that binding of RanGTP and NES peptide and/or SPN1 reduces the shape to a more compact conformation.

Single Particle EM Structures of Human CRM1 in the Free Form

We next addressed the question whether the different conformations of free hsCRM1 can also be seen on a single molecule level in a noncrystalline environment. For this purpose, hsCRM1 was subjected to the GraFix approach—a method allowing the stabilization of different structural populations that exist in solution—and subsequent single-particle EM analysis (Figures S6 and S7). As expected, the human sample showed much higher flexibility when fixed at 4°C compared to our previous study on the *C. thermophilum* CRM1 (Monecke et al., 2013). Thus, to reduce the number of conformations, this stabilization was performed at −10°C. As also seen for ctCRM1, free hsCRM1 occurs in two different and clearly distinct conformations. However, while approximately two thirds of ctCRM1 adopt an extended and pitched superhelical conformation, about half of the human particles (19,254 of 42,108) classified to this shape (Figure 3), similar to that seen in the crystal structure of free ctCRM1. The other conformer, represented by the remaining half of the particles, resembles the shape of a distorted toroid, reminiscent of the CRM1 conformation observed in various binary and ternary complexes. Interestingly, in contrast to the *C. thermophilum* homolog, resampling methods allowed us to predict a large number of subpopulations for the compact conformer, which could not be separated further (Figure S7).

The observation of the high conformational flexibility of free hsCRM1 in the EM prompted us to reinvestigate the results of free CRM1 obtained by SAXS. As mentioned previously, neither the extended nor the compact form of CRM1 fit the SAXS data well. Moreover, the R_g determined experimentally lies within the range between the calculated R_g of the extended and

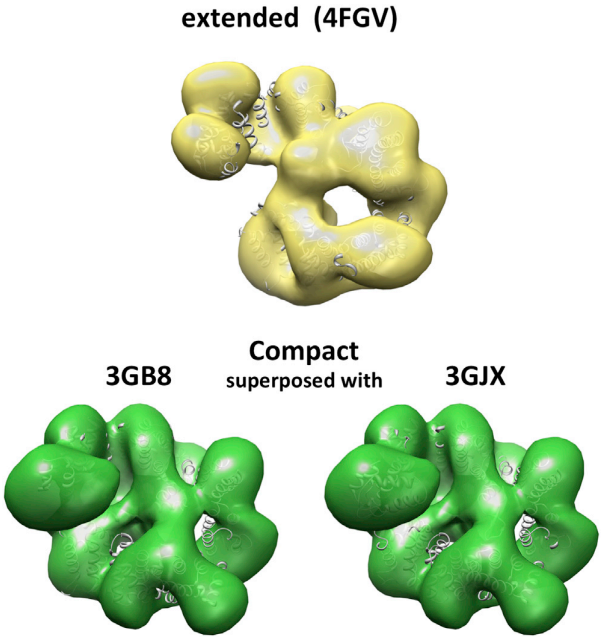


Figure 3. Electron Microscopy Analysis of Free *Homo sapiens* CRM1
EM models of the compact (green) and the extended conformation (yellow) of free *hs*CRM1 (see also Figures S6 and S7). The crystal structures of free *ct*CRM1 (4FGV) and CRM1 derived from the complex structures with SPN1 (3GB8) or the ternary complex with RanGTP and SPN1 (3GJX) are fitted to the envelope models of the EM structures as indicated.
See also Table S3.

compact conformations of CRM1, suggesting a mixture of these two conformations. The best fit for the experimental data of free CRM1 from human and mouse could therefore be obtained with a mixed population using a ratio of roughly 40:60 between extended and compact structures of CRM1 (4FGV/3GJX; Table S3). Please note that only these crystal structures were used because they include 98% of all atoms.

Taken together, the EM results show that free CRM1 in solution exists in open extended, superhelical conformations alongside the compact circular conformations. Whether the observed compact conformations are fully compact as in the ternary complex structure or represent the almost compact conformations as seen for the CRM1-SPN1 complex, cannot be answered unambiguously. The fact that no extended structure was observed in the 100-ns MD simulations indicates that the different conformations are separated by considerable energy barriers.

MD Simulation: Toward an Open CRM1 Structure

To better understand the nature of the forces that oppose the opening of the compact CRM1 and the increase of the superhelical pitch, we focused on two sites of interest both residing within CRM1, i.e., the AL and the contact site between the N- and C-terminal regions, including the C-terminal helix 21B (Dong et al., 2009b; Koyama and Matsuura, 2010; Monecke et al., 2009). The crystal structures suggest, that, on one hand, the AL tends to stabilize a compact CRM1 conformation when engulfing Ran (Figure 1A). This conformation is rearranged when RanBP1 is bound (as in 3M1I) toward a "flipped back" conformation, and remains in this state in the more extended conformations (as in 4FGV, 4HZK, and 3VYC). Helix 21B, on the other hand, adopts two different conformations in the available crystal structures. In the compact form, it is arranged in the HEAT repeat-like "parallel" fashion and located at the outside of CRM1 (3GJX). In contrast, in the other conformations it spans the central opening of CRM1 ("bridging"), contacting residues in the region forming the NES-binding cleft (3GB8, 4FGV, 4HZK, and 3VYC; Figure 1A). The major differences between the extended form and an observed intermediate, the almost compact conformation, are the number of contacts between the N- and C-terminal regions and the fact that the C-terminal acidic patch is in contact with the HEAT repeats that line the NES-binding cleft only in the extended conformation.

As a first step, the role of the N- and C-terminal interactions in maintaining a toroidal conformation was investigated in the presence and absence of the AL. In force-probe MD simulations, both the N- and C-terminal regions of CRM1 were subjected to pulling potentials acting in opposite directions. The forces were applied close to the interface where the N- and C-terminal regions contact each other to form a toroid or closed solenoid (Figure S8A). In most simulations, the force led to rupture of the ring-closing contacts without severely perturbing the HEAT repeats (Figure 4A). All simulations resulted in extended, superhelical structures with a high flexibility and a varying degree of pitch within less than 1 ns. These global conformations are quite similar to the open conformations of superhelical KAPs, such as the prototypic solenoid Impβ. Rupture of the toroid interface is associated with a peak in the force curves (Figure 4A), seen here at 0.4 ns simulation time. To test whether enforced opening of CRM1 is reversible, we performed relaxation simulations, allowing the extended conformational states of CRM1 to evolve freely (Figure 4C). Indeed, within 10 ns, CRM1 recovered a ring structure after release of the pulling force, as indicated by a decrease of both R_g and the rmsd relative to the compact conformation (Figure 4B). An overlay of the recovered conformation with the initial CRM1 ring shows their high structural similarity (Figure 4C). A notable exception is the exact pattern of close contacts at the interface between the N- and C-terminal regions.

In summary, these simulations show that CRM1 can be brought into an elongated, superhelical conformation similar to Impβ when the contact between the N- and C-terminal regions is ruptured by external mechanical strain. The extended conformation of CRM1 shows the major hallmarks of an α solenoid, i.e., high overall flexibility under simultaneous stability of the secondary structure elements. The remarkably high transition rates observed for returning to its original equilibrium conformation are similar to those seen for the global conformational changes of Impβ (Kappel et al., 2010). They thus appear to be a general feature of nuclear transport receptors.

To further characterize the driving forces for connecting the N- and C-terminal regions and for stabilizing the connection, we tested whether the mechanical properties of CRM1 after a cycle of pulling and relaxation are similar to those of the initial structure. Stretching simulations were repeated on relaxed CRM1 structures with pulling potentials acting on the terminal sections.

Structure
Structural Insight into Mammalian CRM1 Cooperativity

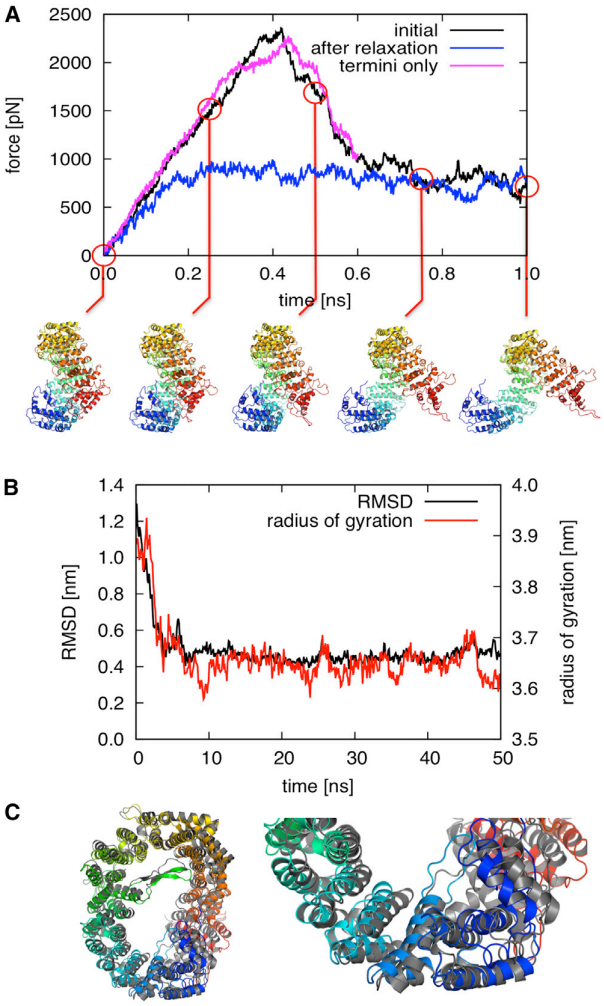


Figure 4. CRM1 Stretching and Relaxation

(A) Force profiles obtained from simulations with a probe velocity of v = 5 m/s: initial stretching (black), stretching after relaxation (blue), and stretching of a structure containing only the terminal regions (magenta). The red circles denote the times the snapshots shown below were taken. Colors are in rainbow progressing from N terminus (blue) to C terminus (red).

(B) Backbone rmsd with respect to the initial structure (solid black lines) and R_g (solid red line) of CRM1 during relaxation.

(C) The left panel shows an overlay of the initial CRM1 structure (gray) and a structure after 50 ns of relaxation. The right panel shows a close-up of the region connecting the termini. Structure and colors as in (B).

See Figure S8 for experimental setup and additional information.

In contrast to the initial simulations, repeated stretching did no longer lead to pronounced force peaks, i.e., a much lower force was now required to separate the terminal regions of relaxed CRM1 (Figure 4A). This lack of interaction forces might either be due to a perturbation of the global elastic properties of CRM1 caused by the opening/closing cycle, i.e., a change in the microscopic interaction pattern within and between all HEAT repeats, or, alternatively, to the loss of important interactions at the interface between the terminal regions.

To differentiate between these two possibilities, additional force probe simulations with only N- and C-terminal fragments of CRM1 were conducted (residues A12–V274 and I815–S1055, in the absence of in-lying HEAT repeats; Figure S8B). The observed force peaks and required force for separating the terminal fragments is nearly identical to that needed to rupture these contacts in full CRM1 (Figure 4A). This suggests that the main contribution to ring closure comes from the interfacial contacts between the N- and C-terminal regions rather than from strain within the body of CRM1. Indeed, because the forces needed to rupture the terminal interface are larger than those seen for stretching the protein, the terminal interactions might even serve to maintain mechanical strain and, thereby, store energy within the array of HEAT repeats. Overall, the force probe simulations suggest that the arrangement of HEAT repeats is compatible with both extended and compact conformations of the free CRM1. The latter is stabilized by specific interactions between the terminal regions. After stretching and release, these contacts are not fully recovered during the simulations, probably due to the presence of many local minima, separated by high energy barriers.

In contrast to other KAPs, the AL in CRM1 is markedly longer and forms a more rigid structural element consisting of a long β-hairpin. It links the two α helices of H9 and affixes RanGTP to CRM1 like a seatbelt, contacting H12–H15 opposite H9. The observed local rigidity in that region is an intrinsic property of CRM1. We next analyzed whether it directly arises from interactions of the AL by performing simulations on a fragment of CRM1 comprising only the central HEAT repeats including the AL (residues R344–L811; Figure S8B). The structure of this fragment remained stable for 50 ns, as shown by its R_g and structural snapshots (Figures 5A and 5D). Closer analysis revealed that three residues within the AL (D436, E439, and R442) form particularly strong electrostatic interactions to the α helices of H12, H14, and H15. Their role in rigidifying the central CRM1 section was therefore examined further. In simulations of CRM1 charge reversal mutants (triple mutation D436K/E439K/R442E, Figure 5B), in which the interactions of the AL with the opposing face of the CRM1 ring are abolished, the central region showed a significant change in its curvature within 50 ns (Figure 5D). Further simulations, in which these residues were each mutated to alanine, displayed a similar change in shape (Figures 5C and 5D).

In summary, the conformation of CRM1 is regulated by a complex pattern of interactions between successive HEAT repeats, the interface between the terminal regions, the AL, and the C-terminal helix 21B.

MD Analysis of Structural Changes in Ran and NES Binding Sites

Two prominent mechanisms are conceivable to explain how the AL mediates cooperative binding of RanGTP and SPN1. One idea is that the AL in the seatbelt conformation may stabilize the compact conformation, which then shifts the equilibrium at the NES-binding cleft toward a conformation prone for cargo binding. Alternatively, the conformation of the AL might directly determine the conformational state of the NES-binding cleft, thereby coupling the global conformation to the NES-binding site. To test the first idea, we recorded the rupture force in MD

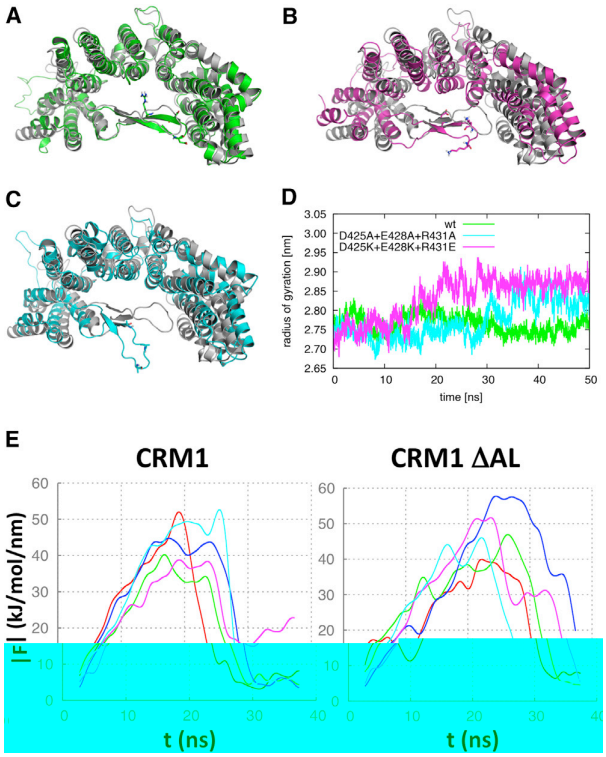


Figure 5. Influence of the AL on CRM1 Conformation

(A–C) Snapshots at the start (gray) and after 50 ns (colored) of each simulation. Key residues are shown in stick and sphere representation. The panels show wt CRM1 (A, green), mutant D436K/E439K/R442E (B, magenta), and mutant D436A/E439A/R442A (C, cyan).

(D) R_g of the WT and the two mutations for each single simulation. Raw data (symbols) and Gaussian filtered data (lines) are shown. Colors as in (A)–(C).

(E) By applying a time-dependent harmonic biasing potential, CRM1 is brought from the compact into the extended conformation. The average over the maximally occurring forces during these force probe simulations, the rupture force, is related to the energetic barrier separating compact and extended conformation. Comparing these rupture forces for WT (left) and AL deletion mutant (right) simulations reveals that the AL does not significantly influence this energetic barrier (see also Table 2).

simulations with an external biasing potential that drives the compact (3GJX) structure toward the extended (4FGV) conformation, both for the wild-type (WT) and an AL deletion mutant (Figure 5E). Remarkably, no significant difference was observed between these variants, and this result was robust under variation of the pulling velocity (Table 2; Figure S8C). These findings suggest that the energy required for the compact-to-extended transition of CRM1 is dominated by the interactions between the C- and N-terminal regions, whereas the AL seems to play a rather minor role. Indeed, closer analysis of the simulations showed that the AL maintains all interactions that stabilize the seatbelt conformation even after this enforced conformational change.

Next we investigated the influence of the AL on the configuration of the NES-binding cleft by carrying out unbiased simulations of WT CRM1 and the AL deletion mutant, both in the presence and absence of RanGTP. Here, the progression of

Table 2. Average Rupture Forces Calculated from Independent Force Probe Simulations for CRM1 WT or AL Deletion Mutant

$F_{rupture}$ kJ/(mol*nm)	3GJX WT	3GJX w/o AL
Initial	45.7 ± 2.6	48.5 ± 2.7
Slower	42.6 ± 2.3	37.0 ± 2.9

The average rupture forces for the force probe simulations in the presence or absence of the AL at two velocities are shown.

the conformational transition of the NES-binding cleft was characterized by projecting its structure onto the difference vector between the compact and extended conformations. In most simulations with bound RanGTP and in the absence of the AL, the NES-binding cleft closed within 10–60 ns (Figure 6B). By contrast, in WT CRM1 the cleft remained open during all five 100-ns simulations (Figure 6A). This observation strongly suggests that the AL mediates the cooperative binding of RanGTP and a cargo protein by stabilizing the open configuration of the NES-binding cleft.

In the absence of RanGTP, the NES-binding cleft of WT CRM1 in the compact conformation adopts an intermediate, semi-open state (Figure 6C). For the AL deletion mutant in the compact conformation without RanGTP, the ensemble of ten trajectories is probably not fully converged, as inferred from the bimodal distribution (Figure 6D). Because several closing but no re-opening events of the NES-binding cleft are seen during the 200-ns simulations, we assume that the kinetics are slowed down, with the closed cleft conformation still favored energetically.

Taken together, these results support the hypothesis that the AL directly determines the NES-binding cleft configuration. In contrast, the AL is unlikely to play a major role in the stabilization of the compact ring-like configuration of CRM1. Our finding that the AL conformation is correlated to the arrangement of the HEAT repeats lining the NES-binding cleft leads us to suggest that RanGTP facilitates cargo uptake by fine-tuning the orientation of the central HEAT repeats and, in particular, the NES-binding cleft between helices 11A and 12A. These results also support a model in which RanBP1 disassembles the complex by causing a rearrangement of the AL (Koyama and Matsuura, 2010), which leads to a shift in the relative free energy of the binding cleft conformations. This in turn decreases the affinity for cargo, resulting in its release and subsequent closure of the NES-binding cleft. Thereby the overall compact conformation of CRM1 is destabilized, facilitating full disassembly of the complex.

To test this idea further, we investigated structural changes among the HEAT repeats upon cargo and Ran binding. These structural units have been shown to be quite rigid, so major overall structural changes predominantly rely on alterations of inter-HEAT repeat interactions (Forwood et al., 2010; Kappel et al., 2010). We monitored the movements of HEAT repeats in unbiased MD trajectories starting from the crystal structure of the ternary complex (3GJX), either complete or after removal of Ran and/or SPN1. Figure 7 shows the backbone rmsd of the 21 individual HEAT repeats. The center of mass (COM) distance of neighboring HEAT repeats is plotted in Figure S9. In all cases, the closed shape of CRM1 remained intact after a simulation time of 100 ns, as reflected by the generally low rmsd with only

Structure
Structural Insight into Mammalian CRM1 Cooperativity

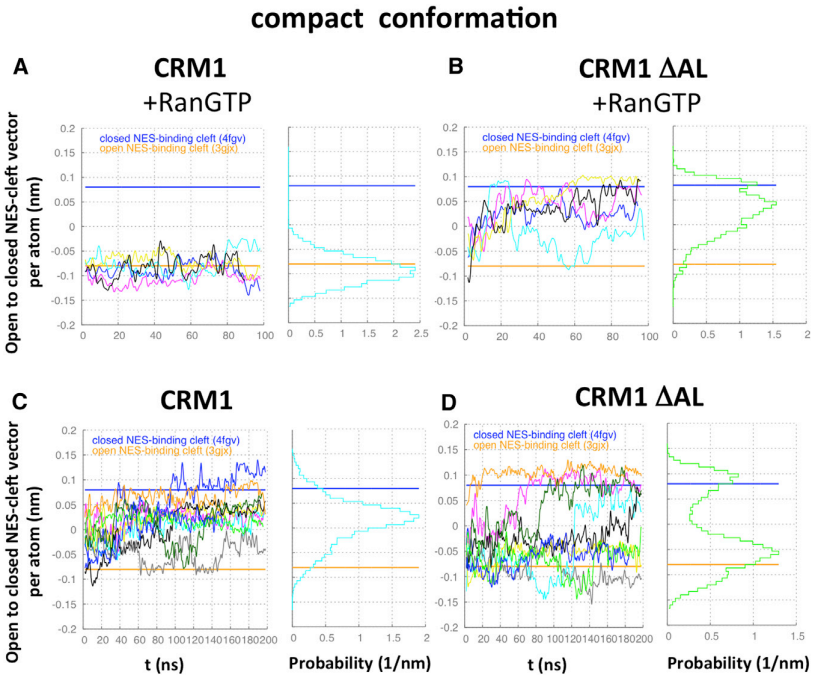


Figure 6. The AL Influences the Conformation of the NES Cleft in the Compact Toroid of CRM1

Projections of unbiased WT and AL deletion mutant simulations onto the vector connecting the open and the closed NES-binding cleft configuration, serving as a reaction coordinate to quantify open/close transitions of the NES-binding cleft. The open NES-binding cleft configuration was taken from the compact CRM1 structure (3GJX), the closed one from the extended CRM1 structure (4FGV). In (A)–(D), the vector coordinate values (per atom) for the open and closed reference structures are shown as horizontal lines. The histograms on the right are constructed from the data of all shown simulations.

(A) In WT simulations and under presence of RanGTP, the NES-binding cleft remains open in all 100-ns simulations.

(B) In AL deletion mutant simulations, spontaneous closure of the NES cleft is seen.

(C) In the absence of RanGTP, the NES-binding cleft adopts an intermediate conformation in WT simulations.

(D) Several closing but no reopening events of the NES-binding cleft are observed in AL deletion mutant simulations in the absence of RanGTP, indicating that the closed conformation is more stable.

the N-terminal HEAT repeats as exception. In light of the EM and X-ray results, this finding implies that either opening of CRM1 is intrinsically slower or that additional factors are required to promote this transition.

In contrast, the three N-terminal HEAT repeats revealed a high degree of flexibility and underwent marked conformational rearrangements during the 100-ns simulations (Figure 7). These HEAT repeats form the main RanGTP binding site (CRIME domain), which is the most highly conserved domain within the Impβ superfamily (Fornerod et al., 1997; Görlich et al., 1997; Petosa et al., 2004). This domain is more flexible than the other HEAT repeats, which suggests a weak binding of RanGTP as shown in biochemical assays (Paraskeva et al., 1999; Petosa et al., 2004). Even with RanGTP bound, the H1 helices show a noticeable degree of conformational fluctuations (Figures 7A and 7D). When RanGTP is removed from the complex, the flexibility of H1 increases further (Figures 7B and 7C), consistent with the CRM1-SPN1 crystal structure.

While changes in flexibility and conformation of the N-terminal region of CRM1 upon RanGTP-binding are clearly reflected in the rmsd, the regions involved in cargo binding seem unaffected by the presence of the binding partners. In the case of SPN1, the binding site is composed of three patches: the NES-binding cleft formed by the outward-oriented α helices of H11 and H12, the intra HEAT loop regions of H12–H14, involved in the interaction with the cap-binding domain, and the binding site for the SPN1 C-terminal region, formed by the α helices of H14–H16. Because the NES-binding cleft is the most important of these patches, the putative changes in H11 and H12 were monitored by recording their COM distance in the simulations (Figure 8A). When an NES is bound in the cleft, the distance remains at 1.7 nm (Figures 8B and S10E), as expected from the X-ray structure (3GJX). This value agrees well with other cargo-bound CRM1 structures (1.59–1.64 nm distance; 3GJX, 3NC0, 3NBY, 3NBZ, and 3GB8; Figure S11). Removal of both RanGTP and SPN1 from the starting model results in a fast decrease of the distance between H11 and H12, indicating a closure of the NES-binding cleft toward a conformation incompatible with NES binding (Figures 8B and S10A). In free CRM1, the distance decreases for all trajectories from 1.7 nm to less than 1.5 nm and as low as 1.3 nm. In all cases, this conformation was attained within the first 50 ns and thereafter remained "closed" (Figure 8B). This finding is in agreement with previous simulations of free ctCRM1 in the extended conformation (Monecke et al., 2013), where the probability to observe the NES-binding cleft in an open conformation was consistently below 20%. When only the NES-bearing cargo, here SPN1, was removed from the complex with RanGTP still bound, larger distance fluctuations were seen; however, in four of five trajectories, the average distance remained within 0.1 nm of those obtained for the ternary complex, and similar to the respective X-ray structure (1.64 nm; 3NC1). The fifth trajectory eventually approached a more closed conformation (Figure S10B and S10C). In contrast, the AL remained in the original seatbelt conformation in all simulations (Figure S1B). We conclude that, although RanGTP is not in direct contact with H11 and H12, RanGTP binding markedly shifts the equilibrium from a closed conformation of the NES-binding cleft toward an open one, capable of binding cargo.

Interestingly, in the X-ray structures of CRM1-RanGTP-RanBP1, the AL is found in a "flipped back" configuration, which might prevent more pronounced changes in the conformation of the NES-binding cleft. In one of the structures, the NES-binding cleft is empty, which is probably why, previously, this AL arrangement was assumed to displace the cargo from the NES-binding cleft and prevent cargo rebinding (3M1I; Koyama

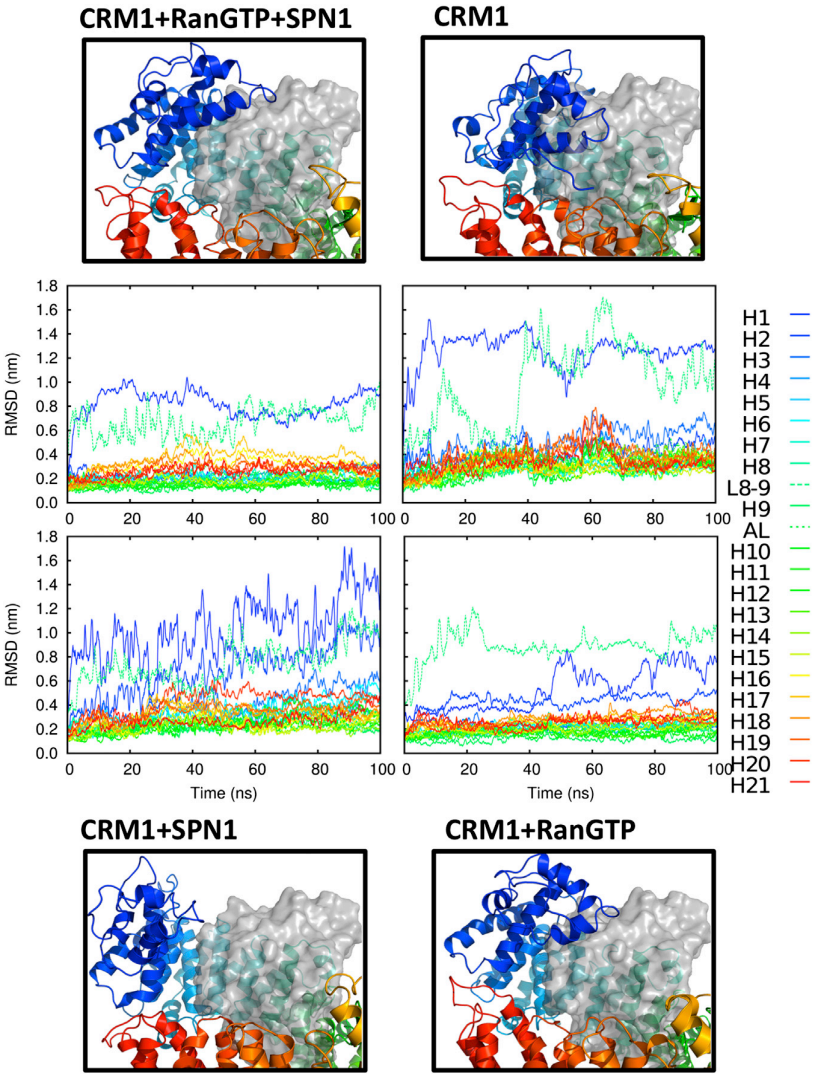


Figure 7. Significant Structural Rearrangements in CRM1 Related to the Respective Binding Partners Are Predominantly Observed within the N-Terminal HEAT Repeats

The spatial changes (C rmsd) of the 21 individual HEAT repeats are plotted over time with CRM1 from the crystal structure 3GJX as reference. The individual HEAT-repeats and relevant loops are labeled according to the color code shown on the right. The most prominent changes within the simulations are observed in the three N-terminal HEAT repeats (see also Figure S9). The respective structures are presented: CRM1-RanGTP-SPN1, CRM1, CRM1-SPN1, and CRM1-RanGTP.

and Matsuura, 2010). Recently, two additional crystal structures of this complex with small inhibitors of nuclear export bound in the NES-binding cleft have been determined (4GPT and 4GMX; Etchin et al., 2013; Lapalombella et al., 2012). All three structures display H11–H12 distances between 1.44 and 1.60 nm, similar to those obtained from our MD simulations. This finding indicates, that despite the binding of RanGTP and RanBP1 and the resulting rearrangements in the HEAT repeats around the NES-binding cleft, the cleft is still flexible enough to accommodate "cargo". In contrast, the three recently published extended conformations of CRM1 exhibit H11A–H12A COM distances between 1.31 and 1.39 nm (4FGV, 4HZK, and 3VYC). Strikingly, the width of the NES-binding cleft increases from the most extended conformation of CRM1 (4FGV) to the least extended one (3VYC). This is in good agreement with the finding that the populations of the NES-binding cleft conformations are closely coupled to the extension of the overall CRM1 structure (Monecke et al., 2013) and could resemble states more or less prone for cargo binding.

Taken together, the MD simulations show that even in the compact toroidal conformation of CRM1, RanGTP binding markedly shifts the equilibrium toward the open conformation of the NES-binding cleft, thus favoring NES binding.

Conclusions

By combining X-ray crystallography, SAXS, single particle-EM, and atomic simulations, we showed that free human and mouse CRM1 are both highly flexible molecules. Free CRM1 can adopt multiple conformations as shown by electron microscopy and indicated by SAXS, ranging from extended conformations without interactions between N- and C-terminal regions to almost compact ones. In ternary complexes, CRM1 is in a compact toroidal conformation, corresponding to that in the known crystal structures of export complexes. Our experimental data extend earlier studies on ctCRM1 and scCRM1 to higher eukaryotes, and show that structural rearrangements are a general property of CRM1.

Our MD simulations confirm the high flexibility of CRM1 and show that CRM1 can reversibly switch from compact to extended conformations without disrupting the array of HEAT repeats. The toroidal shape of CRM1 is mainly stabilized by strong interactions between the N- and C-terminal regions. The exact compact state conformation of CRM1 is determined by an unexpectedly complex interplay of several structural features and their mutual interactions, such as the arrangement of the HEAT repeats, conformation of the AL, and positions of the C-terminal helix 21B and C-terminal acidic patch.

Our simulations strongly suggest that RanGTP binding favors the compact conformation of CRM1. The AL is the internal CRM1 key mediator transmitting the effect of this global conformational change to the NES-binding cleft. These changes shift the equilibrium of the NES-binding cleft from a closed conformation, which is incapable of substrate binding, toward open binding-competent states, thus enabling cooperative binding of both, RanGTP and cargo. These changes also seem

Structure
Structural Insight into Mammalian CRM1 Cooperativity

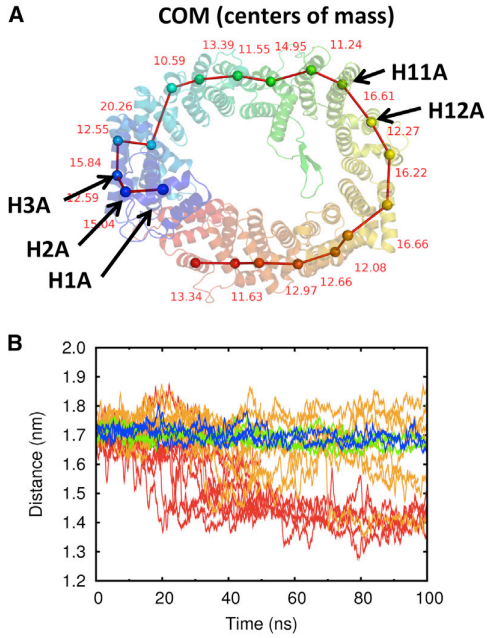


Figure 8. The NES-Binding Cleft Is Stabilized in an Open Conformation by RanGTP

(A) The COM of the α helices of the individual HEAT (colored spheres) repeats were calculated and their distances to neighboring HEATs monitored. Helices referred to in the text are labeled and indicated by arrows.

(B) Time evolution of the COM distance between helices 11A and 12A for the simulations of the ternary complex (blue), CRM1-RanGTP (orange), CRM1-SPN1 (green), and CRM1 alone (red; for snapshots, see Figure S10 and for correlation to known X-ray structures, see Figure S11).

to reduce the free energy barriers that separate the open from the closed state. In this way, binding of RanGTP and cargo protein at two binding sites, separated by a remarkable distance, is coupled both in terms of binding free energies and kinetics, which rationalizes the observed cooperativity in structural terms.

EXPERIMENTAL PROCEDURES

Expression and Purification

CRM1 from *Mus musculus*, RanQ69LGTP 1–180 (referred to as RanGTP in the text) as well as Snurportin1 both from *Homo sapiens* were expressed and purified as described (Monecke et al., 2009). The CRM1-RanGTP-SPN1 complex as well as the CRM1-RanGTP-PKI-NES complex were assembled and purified as described (Güttler et al., 2010; Monecke et al., 2009). Human His_6-CRM1 was expressed in *Escherichia coli* TG1 as described previously (Guan et al., 2000) and purified as described in detail in the Supplemental Experimental Procedures.

Molecular Dynamics Simulations

MD simulations comparing WT and an AL deletion mutant were carried out using GROMACS 4.5 (Hess et al., 2008; Van Der Spoel et al., 2005) with the Amber99sb force field (Hornak et al., 2006) and the SPC/E water model (Berendsen et al., 1987). All other MD simulations were carried out with the GROMACS 4 program package (Van Der Spoel et al., 2005), using the OPLS-AA force field (Friesner et al., 2001; Jorgensen et al., 1996) and the TIP4P water model (Jorgensen et al., 1983). All simulation systems were based on CRM1 as observed in the ternary complex (3GJX).

Small-Angle X-Ray Scattering

The scattering data from solutions of CRM1 alone and in complex were collected on the X33 beamline (EMBL, DORIS III, Hamburg; Blanchet et al., 2012). The data were processed by standard procedures using PRIMUS and Gnom (Svergun, 1992). The low-resolution ab initio shapes were generated using multiple runs of DAMMIF (Franke and Svergun, 2009) averaged by DAMAVER (Volkov and Svergun, 2003) and SUPCOMB (Svergun and Kozin, 2001). A multiphase shape modeling program MONSA (Svergun, 1999) was used for the low-resolution shape analysis of CRM1 in complex. The scattering from the high-resolution models was calculated with CRYSOL (Svergun et al., 1995).

Electron Microscopy Preparation and Image Processing

Purified human CRM1 was prepared and analyzed as described in (Monecke et al., 2013) with the difference that GraFix was run at −10°C. Final three-dimensional models were obtained at a resolution of approximately 20 Å.

ACCESSION NUMBERS

The EMDataBank accession numbers for the structures of CRM1 reported in this paper are EMD-2274 and EMD-5564.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, eleven figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.05.015.

ACKNOWLEDGMENTS

We wish to thank Thomas Güttler and Dirk Görlich for generously sharing with us the mmCRM1-RanGTP-NES complex. We would like to acknowledge the support of the research foundation Deutsche Forschungs Gemeinschaft (SFB860) to R.F. D.H. receives a scholarship from Boehringer Ingelheim Fonds. D.S. and C.B. acknowledge support from the EU e-Infrastructures grant WeNMR, contract number 261572 and the BMBF research grant BIOSCAT, contract number 05K12YE1.

Received: January 10, 2013
Revised: May 7, 2013
Accepted: May 26, 2013
Published: July 11, 2013

REFERENCES

Askjaer, P., Bachi, A., Wilm, M., Bischoff, F.R., Weeks, D.L., Ogniewski, V., Ohno, M., Niehrs, C., Kjems, J., Mattaj, I.W., and Fornerod, M. (1999). RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. Mol. Cell. Biol. *19*, 6276–6285.

Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P. (1987). The missing term in effective pair potentials. J. Phys. Chem. *91*, 6269–6271.

Blanchet, C.E., Zozulya, A.V., Kikhney, A.G., Franke, D., Konarev, P.V., Shang, W., Klaering, R., Robrahn, B., Hermes, C., Cipriani, F., et al. (2012). Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg. J. Appl. Cryst. *45*, 489–495.

Chook, Y.M., and Süel, K.E. (2011). Nuclear import by karyopherin-βs: recognition and inhibition. Biochim. Biophys. Acta *1813*, 1593–1606.

Cook, A.G., and Conti, E. (2010). Nuclear export complexes in the frame. Curr. Opin. Struct. Biol. *20*, 247–252.

Cook, A., Fernandez, E., Lindner, D., Ebert, J., Schlenstedt, G., and Conti, E. (2005). The structure of the nuclear export receptor Cse1 in its cytosolic state reveals a closed conformation incompatible with cargo binding. Mol. Cell *18*, 355–367.

Cook, A.G., Fukuhara, N., Jinek, M., and Conti, E. (2009). Structures of the tRNA export factor in the nuclear and cytosolic states. Nature *461*, 60–65.

Dong, X., Biswas, A., and Chook, Y.M. (2009a). Structural basis for assembly and disassembly of the CRM1 nuclear export complex. Nat. Struct. Mol. Biol. *16*, 558–560.

Dong, X., Biswas, A., Süel, K.E., Jackson, L.K., Martinez, R., Gu, H., and Chook, Y.M. (2009b). Structural basis for leucine-rich nuclear export signal recognition by CRM1. Nature *458*, 1136–1141.

Etchin, J., Sun, Q., Kentsis, A., Farmer, A., Zhang, Z.C., Sanda, T., Mansour, M.R., Barcelo, C., McCauley, D., Kauffman, M., et al. (2013). Antileukemic activity of nuclear export inhibitors that spare normal hematopoietic cells. Leukemia *27*, 66–74.

Fontes, M.R., Teh, T., and Kobe, B. (2000). Structural basis of recognition of monopartite and bipartite nuclear localization sequences by mammalian importin-alpha. J. Mol. Biol. *297*, 1183–1194.

Fornerod, M., van Deursen, J., van Baal, S., Reynolds, A., Davis, D., Murti, K.G., Fransen, J., and Grosveld, G. (1997). The human homologue of yeast CRM1 is in a dynamic subcomplex with CAN/Nup214 and a novel nuclear pore component Nup88. EMBO J. *16*, 807–816.

Forwood, J.K., Lange, A., Zachariae, U., Marfori, M., Preast, C., Grubmüller, H., Stewart, M., Corbett, A.H., and Kobe, B. (2010). Quantitative structural analysis of importin-β flexibility: paradigm for solenoid protein structures. Structure *18*, 1171–1183.

Fox, A.M., Ciziene, D., McLaughlin, S.H., and Stewart, M. (2011). Electrostatic interactions involving the extreme C terminus of nuclear export factor CRM1 modulate its affinity for cargo. J. Biol. Chem. *286*, 29325–29335.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342–346.

Friesner, R.A., Kaminski, G.A., Tirado-Rives, J., and Jorgensen, W.L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J. Phys. Chem. B *105*, 6474–6487.

Görlich, D., Dabrowski, M., Bischoff, F.R., Kutay, U., Bork, P., Hartmann, E., Prehn, S., and Izaurralde, E. (1997). A novel class of RanGTP binding proteins. J. Cell Biol. *138*, 65–80.

Guan, T., Kehlenbach, R.H., Schirmer, E.C., Kehlenbach, A., Fan, F., Clurman, B.E., Arnheim, N., and Gerace, L. (2000). Nup50, a nucleoplasmically oriented nucleoporin with a role in nuclear protein export. Mol. Cell. Biol. *20*, 5619–5630.

Güttler, T., Madl, T., Neumann, P., Deichsel, D., Corsini, L., Monecke, T., Ficner, R., Sattler, M., and Görlich, D. (2010). NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1. Nat. Struct. Mol. Biol. *17*, 1367–1376.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J. Chem. Theory Comput. *4*, 435–447.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins *65*, 712–725.

Hutten, S., and Kehlenbach, R.H. (2007). CRM1-mediated nuclear export: to the pore and beyond. Trends Cell Biol. *17*, 193–201.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Physiol. *79*, 926–935.

Jorgensen, W.L., Maxwell, D.S., and TiradoRives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. *118*, 11225–11236.

Kappel, C., Zachariae, U., Dölker, N., and Grubmüller, H. (2010). An unusual hydrophobic core confers extreme flexibility to HEAT repeat proteins. Biophys. J. *99*, 1596–1603.

Kobe, B., Gleichmann, T., Horne, J., Jennings, I.G., Scotney, P.D., and Teh, T. (1999). Turn up the HEAT. Structure *7*, R91–R97.

Koyama, M., and Matsuura, Y. (2010). An allosteric mechanism to displace nuclear export cargo from CRM1 and RanGTP by RanBP1. EMBO J. *29*, 2002–2013.

Lapalombella, R., Sun, Q., Williams, K., Tangeman, L., Jha, S., Zhong, Y., Goettl, V., Mahoney, E., Berglund, C., Gupta, S., et al. (2012). Selective inhibitors of nuclear export show that CRM1/XPO1 is a target in chronic lymphocytic leukemia. Blood *120*, 4621–4634.

Matsuura, Y., and Stewart, M. (2004). Structural basis for the assembly of a nuclear export complex. Nature *432*, 872–877.

Maurer, P., Redd, M., Solsbacher, J., Bischoff, F.R., Greiner, M., Podtelejnikov, A.V., Mann, M., Stade, K., Weis, K., and Schlenstedt, G. (2001). The nuclear export receptor Xpo1p forms distinct complexes with NES transport substrates and the yeast Ran binding protein 1 (Yrb1p). Mol. Biol. Cell *12*, 539–549.

Monecke, T., Güttler, T., Neumann, P., Dickmanns, A., Görlich, D., and Ficner, R. (2009). Crystal structure of the nuclear export receptor CRM1 in complex with Snurportin1 and RanGTP. Science *324*, 1087–1091.

Monecke, T., Haselbach, D., Voß, B., Russek, A., Neumann, P., Thomson, E., Hurt, E., Zachariae, U., Stark, H., Grubmüller, H., et al. (2013). Structural basis for cooperativity of CRM1 export complex formation. Proc. Natl. Acad. Sci. USA *110*, 960–965.

Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y., and Tsukihara, T. (2009). A high-resolution structure of the pre-microRNA nuclear export machinery. Science *326*, 1275–1279.

Paraskeva, E., Izaurralde, E., Bischoff, F.R., Huber, J., Kutay, U., Hartmann, E., Lührmann, R., and Görlich, D. (1999). CRM1-mediated recycling of snurportin 1 to the cytoplasm. J. Cell Biol. *145*, 255–264.

Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., Soler-López, M., Baudin, F., Mattaj, I.W., and Müller, C.W. (2004). Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. Mol. Cell *16*, 761–775.

Radu, A., Blobel, G., and Moore, M.S. (1995). Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. Proc. Natl. Acad. Sci. USA *92*, 1769–1773.

Saito, N., and Matsuura, Y. (2013). A 2.1-Å-resolution crystal structure of unliganded CRM1 reveals the mechanism of autoinhibition. J. Mol. Biol. *425*, 350–364.

Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. *25*, 495–503.

Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. *76*, 2879–2886.

Svergun, D.I., and Kozin, M.B. (2001). Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. *34*, 33–41.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J. Appl. Crystallogr. *28*, 768–773.

Turner, J.G., Dawson, J., and Sullivan, D.M. (2012). Nuclear export of proteins and drug resistance in cancer. Biochem. Pharmacol. *83*, 1021–1032.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., and Berendsen, H.J. (2005). GROMACS: fast, flexible, and free. J. Comput. Chem. *26*, 1701–1718.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. *36*, 860–864.

 ©2013 Elsevier Ltd All rights reserved

9. Conclusion

The two publications (Monecke et al., 2013; Dölker et al., 2013) gave structural and dynamical insights into the properties of CRM1 and the allosteric binding of cargo and RanGTP. In this conclusion we want to focus on the insights gained by our MD simulations on the allostery.

First we addressed the question what stabilises the extended CRM1 conformation in absence of cargo and RanGTP. Testing various mutants and the wild type, it was found that the C-terminal helix is key to keeping the conformation extended by acting like a ratchet, whereas the acidic loop does not play a role here.

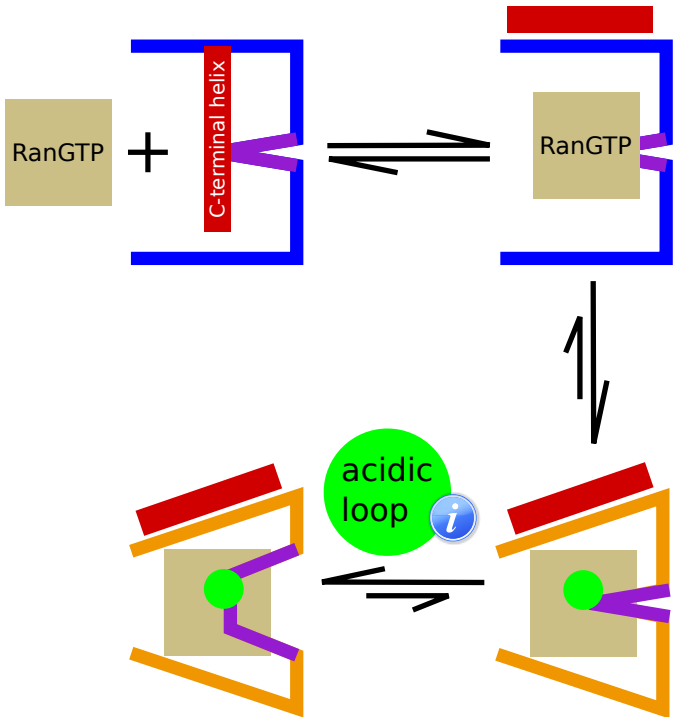
By monitoring the state of the cargo binding site in simulations where a conformational change between extended and compact conformation occurred it was revealed that the probability to find the cargo binding site in the open binding state is significantly increased in the compact conformation. This result is the key to the cooperative binding: Since the NES-cleft has to be in the open state for cargo binding, the affinity for cargo proteins is increased if CRM1 is in the compact conformation.

Based on the x-ray structure of the ternary complex, it was previously speculated that small changes in the ring-like structure, e.g. from a circular to an elliptical conformation were the key to the allostery (Monecke et al., 2009). The ligand-free extended conformation and our MD simulations strongly suggest however that it is a chain of larger conformational changes (i.e. the rearrangement of the C-terminal helix, the subsequent change of the overall conformation from the extended to the compact conformation plus the induced change in the cargo binding site) that is the cause of the allostery.

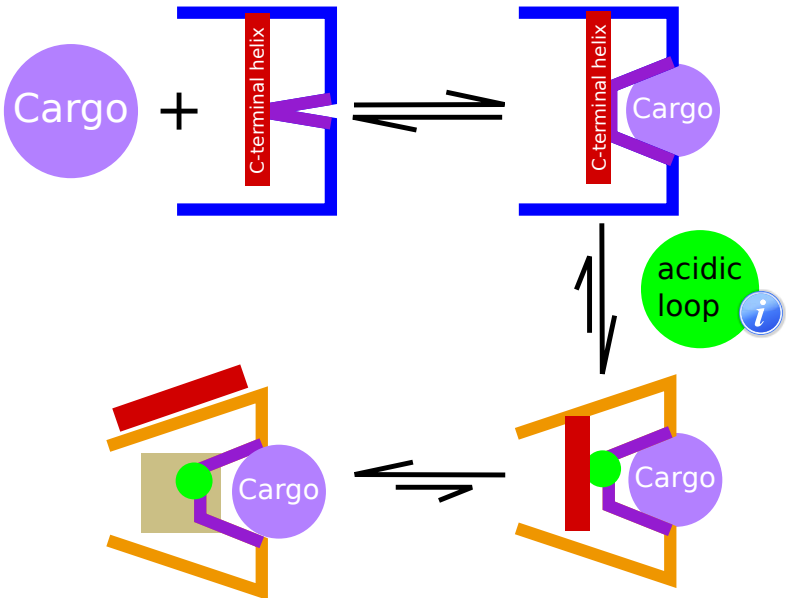
Both in the simulations of wild type CRM1 and the acidic loop deletion mutant the NES cleft stayed open in presence of the acidic loop and spontaneously closed in the absence. This showed that the acidic loop somehow transfers the information of the overall conformation to the cargo binding site and is thus crucial for the allosteric binding. There was no evidence found that the acidic loop contributes stabilises the compact conformation. The second hypothesis, namely that the acidic loop contributes to the allostery by this hypothetical stabilisation, therefore seems less likely, although it cannot be fully excluded.

Putting all these findings together allows us to establish a complete model of the allosteric binding, which is sketched in figure 9.1:

If RanGTP binds to free extended CRM1 (figure 9.1a) at the centre of the ring, the C-terminal helix has to rearrange from the crossing, ratchet like arrangement to the parallel arrangement. This leads to a shift in the free energy landscape of the overall conformation towards the compact conformation. In the compact



(a) RanGTP binds first



(b) Cargo binds first

Figure 9.1.: Model of the mechanism of the allosteric binding.

conformation however, with the help of the acidic loop, the free energy of the cargo binding site is shifted towards the open state, thus increasing the affinity for cargo proteins.

The argumentation holds if a cargo protein binds to free extended CRM1 first as well (figure 9.1b). In this case the cargo binding site has to adopt the open state to allow binding. This change in the cargo binding site then introduces a free energy shift for the overall conformation towards the compact conformation. In the compact conformation however the C-terminal helix cannot be fully extended. Any bending however is accompanied by an energetic cost and this additional cost in free energy reduces the free energy difference to the parallel arrangement of the C-terminal helix, which as a consequence increases the affinity for RanGTP binding.

The model presented here is mainly qualitative. Quantifying the effects of binding on the conformational rearrangements and subsequently on the affinity of the other binding partner in terms of free energies remains subject of future investigations. Such a quantitative analysis in comparison with experimentally determined cooperativity could also help to decide if the mechanism described in this part suffices to explain the allostery or if there are other, until now undiscovered, mechanisms that also contribute to the cooperative binding.

Part IV.

Closing

10. Summary and Outlook

In this work we investigated processes in two systems where binding at a protein is associated with a conformational change in the protein. Our investigations helped to address the questions when, how fast (i.e. with what rate) and why conformational changes occur in the binding process investigated here and shed some light on the effects of these changes.

The process of the first system we investigated is the ligand binding of cAMP at the CNBD of MloK1. Here we decomposed the binding process into substates depending on the protein conformation and ligand presence. We then determined the estimates for the rates between these substates and found that ligand binding occurs as a prerequisite of the conformational change. This conformational change is also determining the effective off-rate.

We furthermore constructed a simple binding funnel model that explains how the ligand reaches the binding site and provides an estimate for the effective on-rate that is in good agreement with the experimentally measured on-rate.

The second system is the exportin CRM1. Here we investigated how the binding of the signal protein RanGTP changes the affinity for cargo proteins such as SPN1 and vice versa even though the binding sites are spatially well separated. Structural information showed that the binding of RanGTP requires the rearrangement of a C-terminal helix. Our simulations revealed that this rearrangement is the key of a global conformational change which in turn causes a local conformational change in the cargo binding site. These coupled conformational changes were found to be the key to the binding allostery.

How transferable are these results to other systems, i.e how prototypical are the systems investigated here? And how transferable are the techniques and models used in this work?

Answering these questions remains a topic for future investigations. We hazard the conjecture that the procedure to define ligand-bound and unbound substates with different conformations and deriving a linear reaction coordinate for the conformational change to calculate the free energy profile for this change will work best if the ensembles of the two conformations are sufficiently well separated in configurational space. Our version of the binding funnel model is probably most accurate for systems were sufficiently small ligands bind in deep binding pockets. A definite answer however requires thorough testing of other systems, including comparisons with experimental data.

Our work on CRM1, i.e. elucidating the reason for the allostery, is bound to be more system specific. In the future and with a drastic increase in computational power it should be possible to calculate PMFs for the conformational changes here

as well, similar as we did for conformational change in the CNBD of MloK1. With these calculations the findings presented in this thesis could be quantified in terms of free energies.

Bibliography

S. L. Altieri, G. M. Clayton, W. R. Silverman, A. O. Olivares, E. M. De La Cruz, L. R. Thomas, and J. H. Morais-Cabral. Structural and energetic analysis of activation by a cyclic nucleotide binding domain. *Journal of Molecular Biology*, 381(3):655–669, Sept. 2008.

J. Aqvist and J. Marelius. The linear interaction energy method for predicting ligand binding free energies. *Combinatorial Chemistry & High Throughput Screening*, 4(8):613–626, Dec. 2001.

P. Askjaer, T. H. Jensen, J. Nilsson, L. Englmeier, and J. Kjems. The specificity of the CRM1-rev nuclear export signal interaction is mediated by RanGTP. *Journal of Biological Chemistry*, 273(50):33414–33422, Dec. 1998.

C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. *The Journal of Physical Chemistry*, 97(40):10269–10280, October 1993.

J. A. Beavo and L. L. Brunton. Cyclic nucleotide research — still expanding after half a century. *Nature Reviews Molecular Cell Biology*, 3(9):710–718, Sept. 2002.

H. J. C. Berendsen. *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*. Cambridge University Press, July 2007.

H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

G. R. Bowman, K. A. Beauchamp, G. Boxer, and V. S. Pande. Progress and challenges in the automated construction of markov state models for full protein systems. *The Journal of Chemical Physics*, 131(12):124101, Sept. 2009.

I. Buch, T. Giorgino, and G. D. Fabritiis. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences*, 108(25):10184–10189, June 2011.

G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

B. M. Carreno and M. Collins. THE b7 FAMILY OF LIGANDS AND ITS RECEPTORS: New pathways for costimulation and inhibition of immune responses. *Annual Review of Immunology*, 20(1):29–53, 2002.

D. A. Case and M. Karplus. Dynamics of ligand binding to heme proteins. *Journal of Molecular Biology*, 132(3):343–368, Aug. 1979.

C.-E. A. Chang, J. Trylska, V. Tozzini, and J. Andrew McCammon. Binding pathways of ligands to HIV-1 protease: Coarse-grained and atomistic simulations. *Chemical Biology & Drug Design*, 69(1):5–13, Jan. 2007.

P.-L. Chiu, M. D. Pagel, J. Evans, H.-T. Chou, X. Zeng, B. Gipson, H. Stahlberg, and C. M. Nimigean. The structure of the prokaryotic cyclic nucleotide-modulated potassium channel MloK1 at 16 Å resolution. *Structure*, 15(9):1053–1064, Sept. 2007.

G. M. Clayton, W. R. Silverman, L. Heginbotham, and J. H. Morais-Cabral. Structural basis of ligand activation in a cyclic nucleotide regulated potassium channel. *Cell*, 119(5):615–627, Nov. 2004.

J. Cohen, A. Arkhipov, R. Braun, and K. Schulten. Imaging the migration pathways for o2, CO, NO, and xe inside myoglobin. *Biophysical Journal*, 91(5): 1844–1857, Sept. 2006.

A. Cook, F. Bono, M. Jinek, and E. Conti. Structural biology of nucleocytoplasmic transport. *Annual Review of Biochemistry*, 76(1):647–671, 2007.

A. G. Cook and E. Conti. Nuclear export complexes in the frame. *Current Opinion in Structural Biology*, 20(2):247–252, Apr. 2010.

P. Csermely, R. Palotai, and R. Nussinov. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends in Biochemical Sciences*, 35(10):539–546, Oct. 2010.

A. Cukkemane, B. Grüter, K. Novak, T. Gensch, W. Bönigk, T. Gerharz, U. B. Kaupp, and R. Seifert. Subunits act independently in a cyclic nucleotide-activated k+ channel. *EMBO reports*, 8(8):749–755, Aug. 2007.

A. Cukkemane, R. Seifert, and U. B. Kaupp. Cooperative and uncooperative cyclic-nucleotide-gated ion channels. *Trends in Biochemical Sciences*, 36(1):55–64, Jan. 2011.

T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An $n \cdot log(n)$ method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12): 10089–10092, 1993.

X. Dong, A. Biswas, K. E. Süel, L. K. Jackson, R. Martinez, H. Gu, and Y. M. Chook. Structural basis for leucine-rich nuclear export signal recognition by CRM1. *Nature*, 458(7242):1136–1141, Apr. 2009.

R. O. Dror, A. C. Pan, D. H. Arlow, D. W. Borhani, P. Maragakis, Y. Shan, H. Xu, and D. E. Shaw. Pathway and mechanism of drug binding to g-protein-coupled receptors. *Proceedings of the National Academy of Sciences of the United States of America*, 108(32):13118–13123, Aug. 2011.

M. Dworkin and K. H. Keller. Solubility and diffusion coefficient of adenosine 3':5'-monophosphate. *Journal of Biological Chemistry*, 252(3):864–865, Feb. 1977.

N. Dölker, C. Blanchet, B. Voß, D. Haselbach, C. Kappel, T. Monecke, D. Svergun, H. Stark, R. Ficner, U. Zachariae, H. Grubmüller, and A. Dickmanns. Structural determinants and mechanism of mammalian CRM1 allostery. *Structure*, 21(8): 1350–1360, Aug. 2013.

D. J. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910–3916, 2005.

P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

K. A. Feenstra, B. Hess, and H. J. C. Berendsen. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry*, 20(8):786–798, 1999.

M. Fornerod, M. Ohno, M. Yoshida, and I. W. Mattaj. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, 90(6):1051–1060, Sept. 1997.

D. Frenkel and B. Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications (Computational Science Series, Vol 1)*. Academic Press, 2nd edition, Nov. 2001. Published: Hardcover.

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe,

P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.

H. Fukunishi, O. Watanabe, and S. Takada. On the hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *The Journal of Chemical Physics*, 116(20):9058–9067, 2002.

C. J. Geyer. Markov chain monte carlo maximum likelihood. *Computing Science and Statistics: Proc. 23rd Symp. Interface*, pages 156–163, 1991.

D. Görlich and U. Kutay. Transport between the cell nucleus and the cytoplasm. *Annual Review of Cell and Developmental Biology*, 15(1):607–660, 1999.

T. Güttler, T. Madl, P. Neumann, D. Deichsel, L. Corsini, T. Monecke, R. Ficner, M. Sattler, and D. Görlich. NES consensus redefined by structures of PKI-type and rev-type nuclear export signals bound to CRM1. *Nature Structural & Molecular Biology*, 17(11):1367–1376, Nov. 2010.

N. C. Handy and A. M. Lee. The adiabatic approximation. *Chemical Physics Letters*, 252(5–6):425–430, Apr. 1996.

T. Hansson, J. Marelius, and J. Åqvist. Ligand binding affinity prediction by linear interaction energy methods. *Journal of Computer-Aided Molecular Design*, 12 (1):27–35, Jan. 1998.

B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, Sept. 1997.

B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, Mar. 2008.

V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, 65(3):712–725, November 2006.

W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

U. B. Kaupp and R. Seifert. Cyclic nucleotide-gated ion channels. *Physiological Reviews*, 82(3):769–824, Jan. 2002.

J. G. Kirkwood. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics*, 3(5):300–313, 1935.

J. Kowal, M. Chami, P. Baumgartner, M. Arheit, P.-L. Chiu, M. Rangl, S. Scheuring, G. F. Schröder, C. M. Nimigean, and H. Stahlberg. Ligand-induced structural changes in the cyclic nucleotide-modulated potassium channel MloK1. *Nature Communications*, 5, Jan. 2014.

K. Kristiansen. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of g-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacology & Therapeutics*, 103(1):21–80, July 2004.

S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

M. W. Mahoney and W. L. Jorgensen. Diffusion constant of the TIP5p model of liquid water. *The Journal of Chemical Physics*, 114(1):363–366, Jan. 2001.

P. Mark and L. Nilsson. Structure and dynamics of the TIP3p, SPC, and SPC/e water models at 298 k. *The Journal of Physical Chemistry A*, 105(43):9954–9960, Nov. 2001.

J. Meyer, Carl D. The role of the group generalized inverse in the theory of finite markov chains. *SIAM Review*, 17(3):pp. 443–464, 1975.

T. Monecke, T. Güttler, P. Neumann, A. Dickmanns, D. Görlich, and R. Ficner. Crystal structure of the nuclear export receptor CRM1 in complex with snurportin1 and RanGTP. *Science*, 324(5930):1087–1091, May 2009.

T. Monecke, D. Haselbach, B. Voß, A. Russek, P. Neumann, E. Thomson, E. Hurt, U. Zachariae, H. Stark, H. Grubmüller, A. Dickmanns, and R. Ficner. Structural basis for cooperativity of CRM1 export complex formation. *Proceedings of the National Academy of Sciences*, 110(3):960–965, Jan. 2013.

J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 1965.

R. P. Newton and C. J. Smith. Cyclic nucleotides. *Phytochemistry*, 65(17):2423–2437, Sept. 2004.

C. M. Nimigean and M. D. Pagel. Ligand binding and activation in a prokaryotic cyclic nucleotide-modulated channel. *Journal of Molecular Biology*, 371(5):1325–1337, Aug. 2007.

C. M. Nimigean, T. Shane, and C. Miller. A cyclic nucleotide modulated prokaryotic k+ channel. *The Journal of General Physiology*, 124(3):203–210, Sept. 2004.

I. H. Pastan, G. S. Johnson, and W. B. Anderson. Role of cyclic nucleotides in growth control. *Annual Review of Biochemistry*, 44(1):491–522, 1975.

M. Patra and M. Karttunen. Systematic comparison of force fields for microscopic simulations of NaCl in aqueous solutions: Diffusion, free energy of hydration, and structural properties. *Journal of Computational Chemistry*, 25(5):678–689, Apr. 2004.

C. Petosa, G. Schoehn, P. Askjaer, U. Bauer, M. Moulin, U. Steuerwald, M. Soler-López, F. Baudin, I. W. Mattaj, and C. W. Müller. Architecture of CRM1/exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Molecular Cell*, 16(5):761–775, Dec. 2004.

S. Peuker, A. Cukkemane, M. Held, F. Noé, U. B. Kaupp, and R. Seifert. Kinetics of ligand-receptor interaction reveals an induced-fit mode of binding in a cyclic nucleotide-activated protein. *Biophysical Journal*, 104(1):63–74, Jan. 2013.

J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte, and F. Noé. Markov models of molecular kinetics: generation and validation. *The Journal of Chemical Physics*, 134(17):174105, May 2011.

S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. v. d. Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, Apr. 2013.

S. Schünke, M. Stoldt, K. Novak, U. B. Kaupp, and D. Willbold. Solution structure of the mesorhizobium loti k1 channel cyclic nucleotide-binding domain in complex with cAMP. *EMBO reports*, 10(7):729–735, July 2009.

S. Schünke, M. Stoldt, J. Lecher, U. B. Kaupp, and D. Willbold. Structural insights into conformational changes of a cyclic nucleotide-binding domain in solution from mesorhizobium loti k1 channel. *Proceedings of the National Academy of Sciences*, 108(15):6121–6126, Apr. 2011.

Y. Shan, E. T. Kim, M. P. Eastwood, R. O. Dror, M. A. Seeliger, and D. E. Shaw. How does a drug molecule find its target binding site? *Journal of the American Chemical Society*, 133(24):9181–9183, June 2011.

M. v. Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift fuer physikalische Chemie*, 92:129–168, 1917.

T. A. Springer. Adhesion receptors of the immune system. *Nature*, 346(6283): 425–434, Aug. 1990.

K. Stade, C. S. Ford, C. Guthrie, and K. Weis. Exportin 1 (crm1p) is an essential nuclear export factor. *Cell*, 90(6):1041–1050, Sept. 1997.

R. H. Swendsen and J.-S. Wang. Replica monte carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57(21):2607–2609, Nov 1986.

W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, Jan. 1982.

R. D. Taylor, P. J. Jewsbury, and J. W. Essex. A review of protein-small molecule docking methods. *Journal of Computer-Aided Molecular Design*, 16(3):151–166, Mar. 2002.

G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, Dec. 2005.

D. van der Spool, E. Lindahl, B. Hess, C. Kurzner, A. R. van Buuren, E. Apol, M. P. J, A. L. Sijbers, K. A. Feenstra, R. v. Drunen, and H. J. Berendsen. *Gromacs User Manual version 4.0.* www.gromcas.org, 2005.

L. Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159(1):98, Jul 1967.

J. Wang and G. M. Verkhivker. Energy landscape theory, funnels, specificity, and optimal criterion of biomolecular binding. *Physical Review Letters*, 90(18): 188101, May 2003.

J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, July 2004.

J. Wang, W. Wang, P. A. Kollman, and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25(2):247 – 260, 2006.

M. Weber. *A Subspace Approach to Molecular Markov State Models via a New Infinitesimal Generator.* Habilitation, Jan. 2012.

H.-J. Woo and B. Roux. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6825–6830, May 2005.

E. Yuriev, M. Agostino, and P. A. Ramsland. Challenges and advances in computational docking: 2009 in review. *Journal of Molecular Recognition*, 24(2): 149–164, Mar. 2011.

J. Åqvist, V. B. Luzhkov, and B. O. Brandsdal. Ligand binding affinities from MD simulations. *Accounts of Chemical Research*, 35(6):358–365, June 2002.

A. Šali and T. L. Blundell. Comparative protein modelling by satisfaction of spatial restraints. *Journal of Molecular Biology*, 234(3):779 – 815, 1993.

K. Šolc and W. H. Stockmayer. Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. i. general theory. *The Journal of Chemical Physics*, 54(7):2981–2988, Apr. 1971.

K. Šolc and W. H. Stockmayer. Kinetics of diffusion-controlled reaction between chemically asymmetric molecules. II. approximate steady-state solution. *International Journal of Chemical Kinetics*, 5(5):733–752, Sept. 1973.

Acknowledgements

First I would like to thank my supervisor Helmut Grubmüller for raising my interest in theoretical and computational biophysics, for his guidance during my work on this thesis and for many fruitful, insightful, exciting and motivating discussions from which I think I gained a tremendous amount of knowledge.

I would like to thank my thesis committee members, Marian Benatti and Jörg Enderlein who always found time for regular meetings.

I would like to thank our to thank our experimental collaborators on the work on the ligand binding at MloK1, Sebastian Peuker, Reinhard Seifert and Benjamin Kaupp for sharing their deep insights in multiple fruitful discussions which proved crucial to the success of the project.

I also would like to thank our experimental collaborators during the work on CRM1, Thomas Monecke, Achim Dickmanns and Ralf Ficner. The close cooperation was a prime example for me how combinations of different experimental and computational techniques provide insights that are hard to obtain otherwise. I also want to thank Ulrich Zachariae who introduced me to the project in a very short time and with whom I had many fruitful discussion.

I also would like to thank all my colleagues and former colleagues in the Department for Theoretical and Computational Biophysics at the MPI for Biophysical Chemistry. I considered the constant exchange of knowledge on science, techniques and non-work related issues of high importance for myself and my work. I want to thank Eveline Heinemann whose organisational capabilities and efficiency always astonished me and who solved most of my organisational issues. I want to thank our system administrators Martin Fechner and Ansgar Esztermann who made the IT run so smoothly one rarely had to worry about it.

Last but not least I want to thank all the unknown people who helped unintentionally and without knowing it: The authors of endless free and open source software packages that I used in my daily work, the people who voluntarily contributed to Wikipedia which often served as a first point for looking up unknown things, and the people out their who answered so many questions on stackoverflow and the likes and who thereby ensured that the answers to many of my programming questions were always at my fingertips.